(19) World Intellectual Property Organization International Bureau



- 1000 Filodox II Kilik Bolica III 100 Folica Kilib Bolica III 100 Folica III 100 Folica III 100 Folica III 1

(43) International Publication Date 17 April 2003 (17.04.2003)

PCT

(10) International Publication Number WO 03/031588 A2

(51) International Patent Classification7:

- -

- (21) International Application Number: PCT/US02/32512
- (22) International Filing Date: 10 October 2002 (10.10.2002)
- (25) Filing Language:

English

C12N

(26) Publication Language:

English

(30) Priority Data:

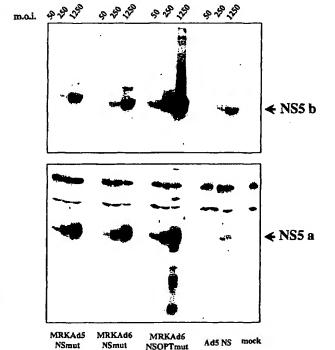
60/328,655 60/363,774 11 October 2001 (11.10.2001) US 13 March 2002 (13.03.2002) US

(71) Applicants (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI, S.P.A. [IT/IT]; VIA PONTINA KM. 30.600, I-00040 POMEZIA (IT).

- (72) Inventors; and
- (75) Inventors, and
 (75) Inventors, Applicants (for US only): EMINI, Emilio,
 A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ
 07065-0907 (US). KASLOW, David, C. [US/US]; 126
 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).
 BETT, Andrew, J. [CA/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). SHIVER, John,
 W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ
 07065-0907 (US). NICOSIA, Alfredo [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). LAHM, Armin
 [DE/IT]; Via Pontina KM. 30.600, I-00040 Pomezia
 (IT). LUZZAGO, Alessandra [IT/IT]; Via Pontina KM.
 30.600, I-00040 Pomezia (IT). CORTESE, Riccardo
 [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT).
 COLLOCA; Stefano [IT/IT]; Via Pontina KM. 30.600,
 I-00040 Pomezia (IT).
- (74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

[Continued on next page]

(54) Title: HEPATITIS C VIRUS VACCINE



(57) Abstract: The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

WO 03/031588 A2



- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CII, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GII, GM, IIR, IIU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PII, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

European patent (AT, BE, BG, CII, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

 without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE OF THE INVENTION HEPATITIS C VIRUS VACCINE

RELATED APPLICATIONS

5

10

15

20

25

30

The present application claims priority to provisional applications U.S. Serial No. 60/363,774, filed March 13, 2002, and U.S. Serial No. 60/328,655, filed October 11, 2001, each of which are hereby incorporated by reference herein.

BACKGROUND OF THE INVENTION

The references cited in the present application are not admitted to be prior art to the claimed invention.

About 3% of the world's population are infected with the Hepatitis C virus (HCV). (Wasley et al., Semin. Liver Dis. 20, 1-16, 2000.) Exposure to HCV results in an overt acute disease in a small percentage of cases, while in most instances the virus establishes a chronic infection causing liver inflammation and slowly progresses into liver failure and cirrhosis. (Iwarson, FEMS Microbiol. Rev. 14, 201-204, 1994.) In addition, epidemiological surveys indicate an important role of HCV in the pathogenesis of hepatocellular carcinoma. (Kew, FEMS Microbiol. Rev. 14, 211-220, 1994, Alter, Blood 85, 1681-1695, 1995.)

Prior to the implementation of routine blood screening for HCV in 1992, most infections were contracted by inadvertent exposure to contaminated blood, blood products or transplanted organs. In those areas where blood screening of HCV is carried out, HCV is primarily contracted through direct percutaneous exposure to infected blood, *i.e.*, intravenous drug use. Less frequent methods of transmission include perinatal exposure, hemodialysis, and sexual contact with an HCV infected person. (Alter *et al.*, *N. Engl. J. Med. 341(8)*, 556-562, 1999, Alter, *J. Hepatol. 31 Suppl.* 88-91, 1999. Semin. Liver. Dis. 201, 1-16, 2000.)

The HCV genome consists of a single strand RNA about 9.5 kb encoding a precursor polyprotein of about 3000 amino acids. (Choo et al., Science 244, 362-364, 1989, Choo et al., Science 244, 359-362, 1989, Takamizawa et al., J. Virol. 65, 1105-1113, 1991.) The HCV polyprotein contains the viral proteins in the order: C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B.

Individual viral proteins are produced by proteolysis of the HCV polyprotein. Host cell proteases release the putative structural proteins C, E1, E2, and

p7, and create the N-terminus of NS2 at amino acid 810. (Mizushima et al., J. Virol. 68, 2731-2734, 1994, Hijikata et al., P.N.A.S. USA 90, 10773-10777, 1993.)

The non-structural proteins NS3, NS4A, NS4B, NS5A and NS5B presumably form the virus replication machinery and are released from the polyprotein. A zinc-dependent protease associated with NS2 and the N-terminus of NS3 is responsible for cleavage between NS2 and NS3. (Grakoui et al., J. Virol. 67, 1385-1395, 1993, Hijikata et al., P.N.A.S. USA 90, 10773-10777, 1993.) A distinct serine protease located in the N-terminal domain of NS3 is responsible for proteolytic cleavages at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A and NS5A/NS5B junctions.

(Bartenschlager et al., J. Virol. 67, 3835-3844, 1993, Grakoui et al., Proc. Natl. Acad. Sci. USA 90, 10583-10587, 1993, Tomei et al., J. Virol. 67, 4017-4026, 1993.)

NS4A provides a cofactor for NS3 activity. (Failla et al., J. Virol. 68, 3753-3760, 1994, De Francesco et al., U.S. Patent No. 5,739,002.)

NS5A is a highly phosphorylated protein conferring interferon resistance. (De Francesco et al., Semin. Liver Dis., 20(1), 69-83, 2000, Pawlotsky, Viral Hepat. Suppl. 1, 47-48, 1999.)

NS5B provides an RNA-dependent RNA polymerase. (De Francesco et al., International Publication Number WO 96/37619, Behrens et al., EMBO 15, 12-22, 1996, Lohmann et al., Virology 249, 108-118, 1998.)

SUMMARY OF THE INVENTION

15

20

25

30

35

The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

A HCV specific CMI response refers to the production of cytotoxic T lymphocytes and T helper cells that recognize an HCV antigen. The CMI response may also include non-HCV specific immune effects.

Preferred nucleic acids encode a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide that is substantially similar to SEQ. ID. NO. 1 and has sufficient protease activity to process itself to produce at least a polypeptide substantially similar to the NS5B region present in SEQ. ID. NO. 1. The produced polypeptide corresponding to NS5B is enzymatically inactive. More preferably, the HCV polypeptide has sufficient

protease activity to produce polypeptides substantially similar to the NS3, NS4A, NS4B, NS5A, and NS5B regions present in SEQ. ID. NO. 1.

Reference to a "substantially similar sequence" indicates an identity of at least about 65% to a reference sequence. Thus, for example, polypeptides having an amino acid sequence substantially similar to SEQ. ID. NO. 1 have an overall amino acid identity of at least about 65% to SEQ. ID. NO. 1.

Polypeptides corresponding to NS3, NS4A, NS4B, NS5A, and NS5B have an amino acid sequence identity of at least about 65% to the corresponding region in SEQ. ID. NO. 1. Such corresponding polypeptides are also referred to herein as NS3, NS4A, NS4B, NS5A, and NS5B polypeptides.

10

15

20

25

30

Thus, a first aspect of the present invention describes a nucleic acid comprising a nucleotide sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The encoded polypeptide has sufficient protease activity to process itself to produce an NS5B polypeptide that is enzymatically inactive.

In a preferred embodiment, the nucleic acid is an expression vector capable of expressing the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide in a desired human cell. Expression inside a human cell has therapeutic applications for actively treating an HCV infection and for prophylactically treating against an HCV infection.

An expression vector contains a nucleotide sequence encoding a polypeptide along with regulatory elements for proper transcription and processing. The regulatory elements that may be present include those naturally associated with the nucleotide sequence encoding the polypeptide and exogenous regulatory elements not naturally associated with the nucleotide sequence. Exogenous regulatory elements such as an exogenous promoter can be useful for expression in a particular host, such as in a human cell. Examples of regulatory elements useful for functional expression include a promoter, a terminator, a ribosome binding site, and a polyadenylation signal.

Another aspect of the present invention describes a nucleic acid comprising a gene expression cassette able to express in a human cell a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The polypeptide can process itself to produce an enzymatically inactive NS5B protein. The gene expression cassette contains at least the following:

a) a promoter transcriptionally coupled to a nucleotide sequence encoding a polypeptide;

5

10

15

20

25

30

35

- b) a 5' ribosome binding site functionally coupled to the nucleotide sequence,
 - c) a terminator joined to the 3' end of the nucleotide sequence, and
- d) a 3' polyadenylation signal functionally coupled to the nucleotide sequence.

Reference to "transcriptionally coupled" indicates that the promoter is positioned such that transcription of the nucleotide sequence can be brought about by RNA polymerase binding at the promoter. Transcriptionally coupled does not require that the sequence being transcribed is adjacent to the promoter.

Reference to "functionally coupled" indicates the ability to mediate an effect on the nucleotide sequence. Functionally coupled does not require that the coupled sequences be adjacent to each other. A 3' polyadenylation signal functionally coupled to the nucleotide sequence facilitates cleavage and polyadenylation of the transcribed RNA. A 5' ribosome binding site functionally coupled to the nucleotide sequence facilitates ribosome binding.

In preferred embodiments the nucleic acid is a DNA plasmid vector or an adenovector suitable for either therapeutic application in treating HCV or as an intermediate in the production of a therapeutic vector. Treating HCV includes actively treating an HCV infection and prophylactically treating against an HCV infection.

Another aspect of the present invention describes an adenovector comprising a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette able to express a polypeptide substantially similar to SEQ. ID. NO. 1 that is produced by a process involving (a) homologous recombination and (b) adenovector rescue. The homologous recombinant step produces an adenovirus genome plasmid. The adenovector rescue step produces the adenovector from the adenogenome plasmid.

Adenovirus genome plasmids described herein contain a recombinant adenovirus genome having a deletion in the E1 region and optionally in the E3 region and a gene expression cassette inserted into one of the deleted regions. The recombinant adenovirus genome is made of regions substantially similar to one or more adenovirus serotypes.

Another aspect of the present invention describes an adenovector consisting of the nucleic acid sequence of SEQ. ID. NO. 4 or a derivative thereof,

wherein said derivative thereof has the HCV polyprotein encoding sequence present in SEQ. ID. NO. 4 replaced with the HCV polyprotein encoding sequence of either SEQ. ID. NO. 3, SEQ. ID. NO. 10 or SEQ. ID. NO. 11.

Another aspect of the present invention describes a cultured recombinant cell comprising a nucleic acid containing a sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The recombinant cell has a variety of uses such as being used to replicate nucleic acid encoding the polypeptide in vector construction methods.

5

10

15

20

25

30

35

Another aspect of the present invention describes a method of making an adenovector comprising a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette able to express a polypeptide substantially similar to SEQ. ID. NO. 1. The method involves the steps of (a) producing an adenovirus genome plasmid containing a recombinant adenovirus genome with deletions in the E1 and E3 regions and a gene expression cassette inserted into one of the deleted regions and (b) rescuing the adenovector from the adenovirus genome plasmid.

Another aspect of the present invention describes a pharmaceutical composition comprising a vector for expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1 and a pharmaceutically acceptable carrier. The vector is suitable for administration and polypeptide expression in a patient.

A "patient" refers to a mammal capable of being infected with HCV. A patient may or may not be infected with HCV. Examples of patients are humans and chimpanzees.

Another aspect of the present invention describes a method of treating a patient comprising the step of administering to the patient an effective amount of a vector expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The vector is suitable for administration and polypeptide expression in the patient.

The patient undergoing treatment may or may not be infected with HCV. For a patient infected with HCV, an effective amount is sufficient to achieve one or more of the following effects: reduce the ability of HCV to replicate, reduce HCV load, increase viral clearance, and increase one or more HCV specific CMI responses. For a patient not infected with HCV, an effective amount is sufficient to achieve one or more of the following: an increased ability to produce one or more components of a HCV specific CMI response to a HCV infection, a reduced

susceptibility to HCV infection, and a reduced ability of the infecting virus to establish persistent infection for chronic disease.

Another aspect of the present invention features a recombinant nucleic acid comprising an Ad6 region and a region not present in Ad6. Reference to "recombinant" nucleic acid indicates the presence of two or more nucleic acid regions not naturally associated with each other. Preferably, the Ad6 recombinant nucleic acid contains Ad6 regions and a gene expression cassette coding for a polypeptide heterologous to Ad6.

Other features and advantages of the present invention are apparent from the additional descriptions provided herein including the different examples. The provided examples illustrate different components and methodology useful in practicing the present invention. The examples do not limit the claimed invention. Based on the present disclosure the skilled artisan can identify and employ other components and methodology useful for practicing the present invention.

15

20

25

30

35

10

5

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B illustrate SEQ. ID. NO. 1.

Figures 2A, 2B, 2C, and 2D illustrate SEQ. ID. NO. 2. SEQ. ID. NO. 2 provides a nucleotide sequence coding for SEQ. ID. NO. 1 along with an optimized internal ribosome entry site and TAAA termination. Nucleotides 1-6 provides an optimized internal ribosome entry site. Nucleotides 7-5961 code for a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide with nucleotides in positions 5137 to 5145 providing a AlaAlaGly sequence in amino acid positions 1711 to 1713 that renders NS5B inactive. Nucleotides 5962-5965 provide a TAAA termination.

Figures 3A, 3B, 3C, and 3D illustrate SEQ. ID. NO. 3. SEQ. ID. NO. 3 is a codon optimized version of SEQ. ID. NO. 2. Nucleotides 7-5961 encode a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

Figures 4A-4M illustrate MRKAd6-NSmut (SEQ. ID. NO. 4). SEQ. ID. NO. 4 is an adenovector containing an expression cassette where the polypeptide of SEQ. ID. NO. 1 is encoded by SEQ. ID. NO. 2. Base pairs 1-450 correspond to the Ad5 bp 1 to 450; base pairs 462 to 1252 correspond to the human CMV promoter; base pairs 1258 to 1267 correspond to the Kozak sequence; base pairs 1264 to 7222 correspond to the NS genes; base pairs 7231 to 7451 correspond to the BGH polyadenylation signal; base pairs 7469 to 9506 correspond to Ad5 base pairs 3511 to 5548; base pairs 9507 to 32121 correspond to Ad6 base pairs 5542 to 28156; base

pairs 32122 to 35117 correspond to Ad6 base pairs 30789 to 33784; and base pairs 35118 to 37089 correspond to Ad5 base pairs 33967 to 35935.

Figures 5A-5O illustrate SEQ. ID. NOs. 5 and 6. SEQ. ID. NO. 5 encodes a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide with an active RNA dependent RNA polymerase. SEQ. ID. NO. 6 provides the amino acid sequence for the polypeptide.

Figures 6A-6C provide the nucleic acid sequence for pV1JnsA (SEQ. ID. NO. 7).

Figures 7A-7N provide the nucleic acid sequence for the Ad6 genome 10 (SEQ. ID. NO. 8).

Figures 8A-8K provide the nucleic acid sequence for the Ad5 genome (SEQ. ID. NO. 9).

Figure 9 illustrates different regions of the Ad6 genome. The linear (35759 bp) ds DNA genome is indicated by two parallel lines and is divided into 100 map units. Transcription units are shown relative to their position and orientation in the genome. Early genes (E1A, E1B, E2A/B, E3 and E4 are indicated by gray arrows. Late genes (L1 to L5), indicated by black arrows, are produced by alternative splicing of a transcript produced from the major late promoter (MLP) and all contain the tripartite leader (1, 2, 3) at their 5' ends. The E1 region is located from approximately 1.0 to 11.5 map units, the E2 region from 75.0 to 11.5 map units, E3 from 76.1 to 86.7 map units, and E4 from 99.5 to 91.2 map units. The major late transcription unit is located between 16.0 and 91.2 map units.

15

20

25

30

35

Figure 10 illustrates homologous recombination to recover pAdE1-E3+ containing Ad6 and Ad5 regions.

Figure 11 illustrates homologous recombinant to recover a pAdE1-E3+ containing Ad6 regions.

Figure 12 illustrates a western blot on whole-cell extracts from 293 cells transfected with plasmid DNA expressing different HCV NS cassettes. Mature NS3 and NS5A products were detected with specific antibodies. "pV1Jns-NS" refers to a pV1JnsA plasmid where a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is encoded by SEQ. ID. NO. 5, and SEQ. ID. NO. 5 is inserted between bases 1881 and 1912 of SEQ. ID. NO. 7. "pV1Jns-NSmut" refers to a pV1JnsA plasmid where SEQ. ID. NO. 2 is inserted between bases 1882 and 1925 of SEQ. ID. NO. 7. "pV1Jns-NSOPTmut" refers to a pV1JnsA plasmid where SEQ. ID. NO. 3 is inserted between bases 1881 and 1905 of SEQ. ID. NO. 7.

Figures 13A and 13B illustrate T cell responses by IFN γ ELIspot induced in C57black6 mice (A) and BalbC mice (B) by two injections of 25 μ g and 50 μ g, respectively, of plasmid DNA encoding the different HCV NS cassettes with Gene Electro-Transfer (GET).

5

10

15

20

25

Figure 14 illustrates protein expression from different adenovectors upon infection of HeLa cells. MRKAd5-NSmut is an adenovector based on an Ad5 sequence (SEQ. ID. NO. 9), where the Ad5 genome has an E1 deletion of base pairs 451 to 3510, an E3 deletion of base pairs 28134 to 30817, and has the NS3-NS4A-NS4B-NS5A-NS5B expression cassette as provided in base pairs 451 to 7468 of SEQ. ID. NO. 4 inserted between positions 450 and 3511. Ad5-NS is an adenovector based on an Ad5 backbone with an E1 deletion of base pairs 342 to 3523, and E3 deletion of base pairs 28134 to 30817 and containing an expression cassette encoding a NS3-NS4A-NS4B-NS5A-NS5B from SEQ. ID. NO. 5. "MRKAd6-NSOPTmut" refers to an adenovector having a modified SEQ. ID. NO. 4 sequence, wherein base pairs 1258 to 7222 of SEQ. ID. NO. 4 is replaced with SEQ. ID. NO. 3.

Figure 15 illustrates T cell responses by IFNγ ELIspot induced in C57black6 mice by two injections of 10⁹ vp of adenovectors containing different HCV non-structural gene cassettes.

Figures 16A-16D illustrate T cell responses by IFN γ ELIspot induced in Rhesus monkeys by one or two injections of 10^{10} vp (A) or 10^{11} vp (B) of adenovectors containing different HCV non-structural gene cassettes.

Figures 17A and 17B illustrates CD8+ T cell responses by IFN γ ICS induced in Rhesus monkeys by two injections of 10^{10} vp (A) or 10^{11} vp (B) of adenovectors encoding the different HCV non-structural gene cassettes.

Figures 18A-18F illustrate T cell responses by bulk CTL assay induced in Rhesus monkeys by two injections of 10¹¹ vp of Ad5-NS (A), MRKAd5-NSmut (B), or MRKAd6-NSmut (C).

Figure 19 illustrates the plasmid pE2.

Figures 20A-D illustrates the partial codon optimized sequence

NSsuboptmut (SEQ. ID. NO. 10). Coding sequence for the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is from base 7 to 5961.

DETAILED DESCRIPTION OF THE INVENTION

10

15

20

25

30

The present invention features Ad6 vectors and nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide that contains an inactive NS5B region. Providing an inactive NS5B region supplies NS5B antigens while reducing the possibility of adverse side effects due to an active viral RNA polymerase. Uses of the featured nucleic acid include use as a vaccine component to introduce into a cell an HCV polypeptide that provides a broad range of antigens for generating a CMI response against HCV, and as an intermediate for producing such a vaccine component.

The adaptive cellular immune response can function to recognize viral antigens in HCV infected cells throughout the body due to the ubiquitous distribution of major histocompatibility complex (MHC) class I and II expression, to induce immunological memory, and to maintain immunological memory. These functions are attributed to antigen-specific CD4+ T helper (Th) and CD8+ cytotoxic T cells (CTL).

Upon activation via their specific T cell receptors, HCV specific Th cells fulfill a variety of immunoregulatory functions, most of them mediated by Th1 and Th2 cytokines. HCV specific Th cells assist in the activation and differentiation of B cells and induction and stimulation of virus-specific cytotoxic T cells. Together with CTL, Th cells may also secrete IFN-γ and TNF-α that inhibit replication and gene expression of several viruses. Additionally, Th cells and CTL, the main effector cells, can induce apoptosis and lysis of virus infected cells.

HCV specific CTL are generated from antigens processed by professional antigen presenting cells (pAPCs). Antigens can be either synthesized within or introduced into pAPCs. Antigen synthesis in a pAPC can be brought about by introducing into the cell an expression cassette encoding the antigen.

A preferred route of nucleic acid vaccine administration is an intramuscular route. Intramuscular administration appears to result in the introduction and expression of nucleic acid into somatic cells and pAPCs. HCV antigens produced in the somatic cells can be transferred to pAPCs for presentation in the context of MHC class I molecules. (Donnelly et al., Annu. Rev. Immunol. 15:617-648, 1997.)

pAPCs process longer length antigens into smaller peptide antigens in the proteasome complex. The antigen is translocated into the endoplasmic reticulum/Golgi complex secretory pathway for association with MHC class I

proteins. CD8+ T lymphocytes recognize antigen associated with class I MHC via the T cell receptor (TCR) and the CD8 cell surface protein.

Using a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide as a vaccine component allows for production of a broad range of antigens capable of generating CMI responses from a single vector. The polypeptide should be able to process itself sufficiently to produce at least a region corresponding to NS5B. Preferred nucleic acids encode an amino acid sequence substantially similar to SEQ. ID. NO. 1 that has sufficient protease activity to process itself to produce individual HCV polypeptides substantially similar to the NS3, NS4A, NS4B, NS5A, and NS5B regions present in SEQ. ID. NO. 1.

5

10

20

25

30

35

A polypeptide substantially similar to SEQ. ID. NO. 1 with sufficient protease activity to process itself in a cell provides the cell with T cell epitopes that are present in several different HCV strains. Protease activity is provided by NS3 and NS3/NS4A proteins digesting the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide at the appropriate cleavage sites to release polypeptides corresponding to NS3, NS4A, NS4B, NS5A, and NS5B. Self- processing of the Met-NS3-NS4A-NS4B-NS5A-NS5B generates polypeptides that approximate naturally occurring HCV polypeptides.

Based on the guidance provided herein a sufficiently strong immune response can be generated to achieve beneficial effects in a patient. The provided guidance includes information concerning HCV sequence selection, vector selection, vector production, combination treatment, and administration.

I. HCV SEQUENCES

A variety of different nucleic acid sequences can be used as a vaccine component to supply a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide to a cell or as an intermediate to produce vaccine components. The starting point for obtaining suitable nucleic acid sequences are preferably naturally occurring NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequences modified to produce an inactive NS5B.

The use of a HCV nucleic acid sequence providing HCV non-structural antigens to generate a CMI response is mentioned by Cho *et al.*, Vaccine 17:1136-1144, 1999, Paliard *et al.*, International Publication Number WO 01/30812 (not admitted to be prior art to the claimed invention), and Coit *et al.*, International Publication Number WO 01/38360 (not admitted to be prior art to the claimed invention). Such references fail to describe, for example, a polypeptide that processes

itself to produce an inactive NS5B, and the particular combinations of HCV sequences and delivery vehicles employed herein.

Modifications to a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequence can be produced by altering the encoding nucleic acid. Alterations can be performed to create deletions, insertions and substitutions.

5

10

15

20

25

30

35

Small modifications can be made in NS5B to produce an inactive polymerase by targeting motifs essentially for replication. Examples of motifs critical for NS5B activity and modifications that can be made to produce an inactive NS5B are described by Lohmann *et al.*, *Journal of Virology 71*:8416-8426, 1997, and Kolykhalov *et al.*, *Journal of Virology 74*:2046-2051, 2000.

Additional factors to take into account when producing modifications to a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide include maintaining the ability to self-process and maintaining T cell antigens. The ability of the HCV polypeptide to process itself is determined to a large extent by a functional NS3 protease. Modifications that maintain NS3 activity protease activity can be obtained by taking into account the NS3 protein, NS4A which serves as a cofactor for NS3, and NS3 protease recognition sites present within the NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

Different modifications can be made to naturally occurring NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequences to produce polypeptides able to elicit a broad range of T cell responses. Factors influencing the ability of a polypeptide to elicit a broad T cell response include the preservation or introduction of HCV specific T cell antigen regions and prevalence of different T cell antigen regions in different HCV isolates.

Numerous examples of naturally occurring HCV isolates are well known in the art. HCV isolates can be classified into the following six major genotypes comprising one or more subtypes: HCV-1/(1a,1b,1c), HCV-2/(2a,2b,2c), HCV-3/(3a,3b,10a), HCV-4/(4a), HCV-5/(5a) and HCV-6/(6a,6b,7b,8b,9a,11a). (Simmonds, J. Gen. Virol., 693-712, 2001.) Examples of particular HCV sequences such as HCV-BK, HCV-J, HCV-N, HCV-H, have been deposited in GenBank and described in various publications. (See, for example, Chamberlain et al., J. Gen. Virol., 1341-1347, 1997.)

HCV T cell antigens can be identified by, for example, empirical experimentation. One way of identifying T cell antigens involves generating a series of overlapping short peptides from a longer length polypeptide and then screening the

PCT/US02/32512 WO 03/031588

T-cell populations from infected patients for positive clones. Positive clones are activated/primed by a particular peptide. Techniques such as IFNy-ELISPOT, IFNy-Intracellular staining and bulk CTL assays can be used to measure peptide activity. Peptides thus identified can be considered to represent T-cell epitopes of the respective pathogen.

HCV T cell antigen regions from different HCV isolates can be introduced into a single sequence by, for example, producing a hybrid NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing regions from two or more naturally occurring sequences. Such a hybrid can contain additional modifications, which preferably do not reduce the ability of the polypeptide to produce an HCV CMI response.

The ability of a modified Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide to process itself and produce a CMI response can be determined using techniques described herein or well known in the art. Such techniques include the use of IFNy-ELISPOT, IFNy-Intracellular staining and bulk CTL assays to measure a HCV specific CMI response.

A. Met-NS3-NS4A-NS4B-NS5A-NS5B Sequences

5

10

15

20

25

SEQ. ID. NO. 1 provides a preferred Met-NS3-NS4A-NS4B-NS5A-NS5B sequence. SEQ. ID. NO. 1 contains a large number of HCV specific T cell antigens that are present in several different HCV isolates. SEQ. ID. NO. 1 is similar to the NS3-NS4A-NS4B-NS5A-NS5B portion of the HCV BK strain nucleotide sequence (GenBank accession number M58335).

In SEQ. ID. NO. 1 anchor positions important for recognition by MHC class I molecules are conserved or represent conservative substitutions for 18 out of 20 known T-cell epitopes in the NS3-NS4A-NS4B-NS5A-NS5B portion of HCV polyproteins. With respect to the remaining two known T-cell epitopes, one has a non-conservative anchor substitution in SEQ. ID. NO. 1 that may still be recognized by a different HLA supertype and one epitope has one anchor residue not conserved. HCV T-cell epitopes are described in Chisari et al., Curr. Top. Microbiol Immunol., 30 242:299-325, 2000, and Lechner et al. J. Exp. Med. 9:1499-1512, 2000.

Differences between the HCV-BK NS3-NS4A-NS4B-NS5A-NS5B nucleotide sequence and SEQ. ID. NO. 1 include the introduction of a methionine at the 5' end and the presence of modified NS5B active site residues in SEQ. ID. NO. 1.

The modification replaces GlyAspAsp with AlaAlaGly (residues 1711-1713) to inactivate NS5B.

5

10

15

20

25

30

35

The encoded HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide preferably has an amino acid sequence substantially similar to SEQ. ID. NO. 1. In different embodiments, the encoded HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide has an amino acid identify to SEQ. ID. NO. 1 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or differs from SEQ. ID. NO. 1 by 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, or 1-20 amino acids.

Amino acid differences between a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide and SEQ. ID. NO. 1 are calculated by determining the minimum number of amino acid modifications in which the two sequences differ. Amino acid modifications can be deletions, additions, substitutions or any combination thereof.

Amino acid sequence identity is determined by methods well known in the art that compare the amino acid sequence of one polypeptide to the amino acid sequence of a second polypeptide and generate a sequence alignment. Amino acid identity is calculated from the alignment by counting the number of aligned residue pairs that have identical amino acids.

Methods for determining sequence identity include those described by Schuler, G.D. in *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Baxevanis, A.D. and Ouelette, B.F.F., eds., John Wiley & Sons, Inc, 2001; Yona, et al., in *Bioinformatics: Sequence, structure and databanks*, Higgins, D. and Taylor, W. eds, Oxford University Press, 2000; and *Bioinformatics: Sequence and Genome Analysis*, Mount, D.W., ed., Cold Spring Harbor Laboratory Press, 2001). Methods to determine amino acid sequence identity are codified in publicly available computer programs such as GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.), BLAST (Altschul et al., J. Mol. Biol. 215(3):403-10, 1990), and FASTA (Pearson, Methods in Enzymology 183:63-98, 1990, R.F. Doolittle, ed.).

In an embodiment of the present invention sequence identity between two polypeptides is determined using the GAP program (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.). GAP uses the alignment method of Needleman and Wunsch. (Needleman, et al., J. Mol. Biol. 48:443-453, 1970.) GAP considers all possible alignments and gap positions between two sequences and creates a global alignment that maximizes the number of matched

residues and minimizes the number and size of gaps. A scoring matrix is used to assign values for symbol matches. In addition, a gap creation penalty and a gap extension penalty are required to limit the insertion of gaps into the alignment. Default program parameters for polypeptide comparisons using GAP are the BLOSUM62 (Henikoff et al., Proc. Natl. Acad. Sci. USA, 89:10915-10919, 1992) amino acid scoring matrix (MATrix=blosum62.cmp), a gap creation parameter (GAPweight=8) and a gap extension pararameter (LENgthweight=2).

More preferred HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptides in addition to being substantially similar to SEQ. ID. NO. 1 across their entire length produce individual NS3, NS4A, NS4B, NS5A and NS5B regions that are substantially similar to the corresponding regions present in SEQ. ID. NO. 1. The corresponding regions in SEQ. ID. NO. 1 are provided as follows: Met-NS3 amino acids 1-632; NS4A amino acids 633-686; NS4B amino acids 687-947; NS5A amino acids 948-1394; and NS5B amino acids 1395-1985.

10

15

20

25

30

35

In different embodiments a NS3, NS4A, NS4B, NS5A and/or NS5B region has an amino acid identity to the corresponding region in SEQ. ID. NO. 1 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99%, or 100%; or an amino acid difference of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, or 1-20 amino acids.

Amino acid modifications to SEQ. ID. NO. 1 preferably maintain all or most of the T-cell antigen regions. Differences in naturally occurring amino acids are due to different amino acid side chains (R groups). An R group affects different properties of the amino acid such as physical size, charge, and hydrophobicity. Amino acids can be divided into different groups as follows: neutral and hydrophobic (alanine, valine, leucine, isoleucine, proline, tyrptophan, phenylalanine, and methionine); neutral and polar (glycine, serine, threonine, tryosine, cysteine, asparagine, and glutamine); basic (lysine, arginine, and histidine); and acidic (aspartic acid and glutamic acid).

Generally, in substituting different amino acids it is preferable to exchange amino acids having similar properties. Substituting different amino acids within a particular group, such as substituting valine for leucine, arginine for lysine, and asparagine for glutamine are good candidates for not causing a change in polypeptide tertiary structure.

Starting with a particular amino acid sequence and the known degeneracy of the genetic code, a large number of different encoding nucleic acid

sequences can be obtained. The degeneracy of the genetic code arises because almost all amino acids are encoded by different combinations of nucleotide triplets or "codons". The translation of a particular codon into a particular amino acid is well known in the art (see, e.g., Lewin GENES IV, p. 119, Oxford University Press, 1990).

5 Amino acids are encoded by codons as follows:

A=Ala=Alanine: codons GCA, GCC, GCG, GCU

C=Cys=Cysteine: codons UGC, UGU

D=Asp=Aspartic acid: codons GAC, GAU

E=Glu=Glutamic acid: codons GAA, GAG

10 F=Phe=Phenylalanine: codons UUC, UUU

G=Gly=Glycine: codons GGA, GGC, GGG, GGU

H=His=Histidine: codons CAC, CAU

I=Ile=Isoleucine: codons AUA, AUC, AUU

K=Lys=Lysine: codons AAA, AAG

15 L=Leu=Leucine: codons UUA, UUG, CUA, CUC, CUG, CUU

M=Met=Methionine: codon AUG

N=Asn=Asparagine: codons AAC, AAU

P=Pro=Proline: codons CCA, CCC, CCG, CCU

Q=Gln=Glutamine: codons CAA, CAG

20 R=Arg=Arginine: codons AGA, AGG, CGA, CGC, CGG, CGU

S=Ser=Serine: codons AGC, AGU, UCA, UCC, UCG, UCU

T=Thr=Threonine: codons ACA, ACC, ACG, ACU

V=Val=Valine: codons GUA, GUC, GUG, GUU

W=Trp=Tryptophan: codon UGG

25 Y=Tyr=Tyrosine: codons UAC, UAU.

Nucleic acid sequences can be optimized in an effort to enhance expression in a host. Factors to be considered include C:G content, preferred codons, and the avoidance of inhibitory secondary structure. These factors can be combined in different ways in an attempt to obtain nucleic acid sequences having enhanced

30 expression in a particular host. (See, for example, Donnelly *et al.*, International Publication Number WO 97/47358.)

The ability of a particular sequence to have enhanced expression in a particular host involves some empirical experimentation. Such experimentation involves measuring expression of a prospective nucleic acid sequence and, if needed,

35 altering the sequence.

B. Encoding Nucleotide Sequences

5

10

15

20

25

30

35

SEQ. ID. NOs. 2 and 3 provide two examples of nucleotide sequences encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B sequence. The coding sequence of SEQ. ID. NO. 2 is similar (99.4% nucleotide sequence identity) to the NS3-NS4A-NS4B-NS5A-NS5B region of the naturally occurring HCV-BK sequence (GenBank accession number M58335). SEQ. ID. NO. 3 is a codon-optimized version of SEQ. ID. NO. 2. SEQ. ID. NOs. 2 and 3 have a nucleotide sequence identity of 78.3%.

Differences between the HCV-BK NS3-NS4A-NS4B-NS5A-NS5B nucleotide (GenBank accession number M58335) and SEQ. ID. NO. 2, include SEQ. ID. NO. 2 having a ribosome binding site, an ATG methionine codon, a region coding for a modified NS5B catalytic domain, a TAAA stop signal and an additional 30 nucleotide differences. The modified catalytic domain codes for a AlaAlaGly (residues 1711-1713) instead of GlyAspAsp to inactivate NS5B.

A nucleotide sequence encoding a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is preferably substantially similar to the SEQ. ID. NO. 2 coding region. In different embodiments, the nucleotide sequence encoding a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide has a nucleotide sequence identify to the SEQ. ID. NO. 2 coding region of at least 65%, at least 75%, at least 85%, at least 95%, at least 99%, or 100%; or differs from SEQ. ID. NO. 2 by 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides.

Nucleotide differences between a sequence coding Met-NS3-NS4A-NS4B-NS5A-NS5B and the SEQ. ID. NO. 2 coding region are calculated by determining the minimum number of nucleotide modifications in which the two sequences differ. Nucleotide modifications can be deletions, additions, substitutions or any combination thereof.

Nucleotide sequence identity is determined by methods well known in the art that compare the nucleotide sequence of one sequence to the nucleotide sequence of a second sequence and generate a sequence alignment. Sequence identity is determined from the alignment by counting the number of aligned positions having identical nucleotides.

Methods for determining nucleotide sequence identity between two polynucleotides include those described by Schuler, in *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Baxevanis, A.D. and Ouelette, B.F.F.,

eds., John Wiley & Sons, Inc, 2001; Yona et al., in Bioinformatics: Sequence, structure and databanks, Higgins, D. and Taylor, W. eds, Oxford University Press, 2000; and Bioinformatics: Sequence and Genome Analysis, Mount, D.W., ed., Cold Spring Harbor Laboratory Press, 2001). Methods to determine nucleotide sequence identity are codified in publicly available computer programs such as GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.), BLAST (Altschul et al., J. Mol. Biol. 215(3):403-10, 1990), and FASTA (Pearson, W.R., Methods in Enzymology 183:63-98, 1990, R.F. Doolittle, ed.).

In an embodiment of the present ivnention, sequence identity between two polynucleotides is determined by application of GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.). GAP uses the alignment method of Needleman and Wunsch. (Needleman et al., J. Mol. Biol. 48:443-453, 1970.) GAP considers all possible alignments and gap positions between two sequences and creates a global alignment that maximizes the number of matched residues and minimizes the number and size of gaps. A scoring matrix is used to assign values for symbol matches. In addition, a gap creation penalty and a gap extension penalty are required to limit the insertion of gaps into the alignment. Default program parameters for polynucleotide comparisons using GAP are the nwsgapdna.cmp scoring matrix (MATrix=nwsgapdna.cmp), a gap creation parameter (GAPweight=50) and a gap extension pararameter (LENgthweight=3).

More preferred HCV Met-NS3-NS4A-NS4B-NS5A-NS5B nucleotide sequences in addition to being substantially similar across its entire length, produce individual NS3, NS4A, NS4B, NS5A and NS5B regions that are substantially similar to the corresponding regions present in SEQ. ID. NO. 2. The corresponding coding regions in SEQ. ID. NO. 2 are provided as follows: Met-NS3, nucleotides 7-1902; NS4A nucleotides 1903-2064; NS4B nucleotides 2065-2847; NS5A nucleotides 2848-4188: NS5B nucleotides 4189-5661.

In different embodiments a NS3, NS4A, NS4B, NS5A and/or NS5B encoding region has a nucleotide sequence identity to the corresponding region in SEQ. ID. NO. 2 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or a nucleotide difference to SEQ. ID. NO. 2 of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides.

10

15

20

25

C. Gene Expression Cassettes

5

10

15

20

25

30

35

A gene expression cassette contains elements needed for polypeptide expression. Reference to "polypeptide" does not provide a size limitation and includes protein. Regulatory elements present in a gene expression cassette generally include: (a) a promoter transcriptionally coupled to a nucleotide sequence encoding the polypeptide, (b) a 5' ribosome binding site functionally coupled to the nucleotide sequence, (c) a terminator joined to the 3' end of the nucleotide sequence, and (d) a 3' polyadenylation signal functionally coupled to the nucleotide sequence. Additional regulatory elements useful for enhancing or regulating gene expression or polypeptide processing may also be present.

Promoters are genetic elements that are recognized by an RNA polymerase and mediate transcription of downstream regions. Preferred promoters are strong promoters that provide for increased levels of transcription. Examples of strong promoters are the immediate early human cytomegalovirus promoter (CMV), and CMV with intron A. (Chapman *et al*, *Nucl. Acids Res.* 19:3979-3986, 1991.) Additional examples of promoters include naturally occurring promoters such as the EF1 alpha promoter, the murine CMV promoter, Rous sarcoma virus promoter, and SV40 early/late promoters and the β-actin promoter; and artificial promoters such as a synthetic muscle specific promoter and a chimeric muscle-specific/CMV promoter (Li *et al.*, *Nat. Biotechnol. 17*:241-245, 1999, Hagstrom *et al.*, *Blood 95*:2536-2542, 2000).

The ribosome binding site is located at or near the initiation codon. Examples of preferred ribosome binding sites include CCACCAUGG, CCGCCAUGG, and ACCAUGG, where AUG is the initiation codon. (Kozak, *Cell* 44:283-292, 1986). Another example of a ribosome binding site is GCCACCAUGG (SEQ..ID. NO. 12).

The polyadenylation signal is responsible for cleaving the transcribed RNA and the addition of a poly (A) tail to the RNA. The polyadenylation signal in higher eukaryotes contains an AAUAAA sequence about 11-30 nucleotides from the polyadenylation addition site. The AAUAAA sequence is involved in signaling RNA cleavage. (Lewin, Genes IV, Oxford University Press, NY, 1990.) The poly (A) tail is important for the mRNA processing.

Polyadenylation signals that can be used as part of a gene expression cassette include the minimal rabbit β -globin polyadenylation signal and the bovine growth hormone polyadenylation (BGH). (Xu et al., Gene 272:149-156, 2001, Post et

al., U.S. Patent U. S. 5,122,458.) Additional examples include the Synthetic Polyadenylation Signal (SPA) and SV40 polyadenylation signal. The SPA sequence is as follows: AAUAAAAGAUCUUUAUUUUCAUUAGAUCUGUGUGUUUUUUUGUGUG (SEQ. ID. NO. 13).

Examples of additional regulatory elements useful for enhancing or regulating gene expression or polypeptide processing that may be present include an enhancer, a leader sequence and an operator. An enhancer region increases transcription. Examples of enhancer regions include the CMV enhancer and the SV40 enhancer. (Hitt et al., Methods in Molecular Genetics 7:13-30, 1995, Xu, et al., Gene 272:149-156, 2001.) An enhancer region can be associated with a promoter.

A leader sequence is an amino acid region on a polypeptide that directs the polypeptide into the proteasome. Nucleic acid encoding the leader sequence is 5' of a structural gene and is transcribed along the structural gene. An example of a leader sequences is tPA.

An operator sequence can be used to regulate gene expression. For example, the Tet operator sequence can be used to repress gene expression.

II. THERAPEUTIC VECTORS

Nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide can be introduced into a patient using vectors suitable for therapeutic administration. Suitable vectors can deliver nucleic acid into a target cell without causing an unacceptable side effect.

Cellular expression is achieved using a gene expression cassette encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide. The gene expression cassette contains regulatory elements for producing and processing a sufficient amount of nucleic acid inside a target cell to achieve a beneficial effect.

Examples of vectors that can be used for therapeutic applications include first and second generation adenovectors, helper dependent adenovectors, adeno-associated viral vectors, retroviral vectors, alpha virus vectors, Venezuelan Equine Encephalitis virus vector, and plasmid vectors. (Hitt, et al., Advances in Pharmacology 40:137-206, 1997, Johnston et al., U.S. Patent No. 6,156,588, and Johnston et al., International Publication Number WO 95/32733.) Preferred vectors for introducing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide into a subject are first generation adenoviral vectors and plasmid DNA vectors.

30

5

10

15

20

A. First Generation Adenovectors

5

10

15

20

25

30

35

First generation adenovector for expressing a gene expression cassette contain the expression cassette in an E1 and optionally E3 deleted recombinant adenovirus genome. The deletion in the E1 region is sufficiently large to remove elements needed for adenoviral replication.

First generation adenovectors for expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide contain a E1 and E3 deleted recombinant adenovirus genome. The deletion in the E1 region is sufficiently large to remove elements needed for adenoviral replication. The combinations of deletions of the E1 and E3 regions are sufficiently large to accommodate a gene expression cassette encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

The adenovirus has a double-stranded linear genome with inverted terminal repeats at both ends. During viral replication, the genome is packaged inside a viral capsid to form a virion. The virus enters its target cell through viral attachment followed by internalization. (Hitt et al., Advances in Pharmacology 40:137-206, 1997.)

Adenovectors can be based on different adenovirus serotypes such as those found in humans or animals. Examples of animal adenoviruses include bovine, porcine, chimp, murine, canine, and avian (CELO). Preferred adenovectors are based on human serotypes, more preferably Group B, C, or D serotypes. Examples of human adenovirus Group B, C, D, or E serotypes include types 2 ("Ad2"), 4 ("Ad4"), 5 ("Ad5"), 6 ("Ad6"), 24 ("Ad24"), 26 ("Ad26"), 34 ("Ad34") and 35 ("Ad35"). Adenovectors can contain regions from a single adenovirus or from two or more adenovirus.

In different embodiments adenovectors are based on Ad5, Ad6, or a combination thereof. Ad5 is described by Chroboczek, et al., J. Virology 186:280-285, 1992. Ad6 is described in Figures 7A-7N. An Ad6 based vector containing Ad5 regions is described in the Example section provided below.

Adenovectors do not need to have their E1 and E3 regions completely removed. Rather, a sufficient amount the E1 region is removed to render the vector replication incompetent in the absence of the E1 proteins being supplied in *trans*; and the E1 deletion or the combination of the E1 and E3 deletions are sufficiently large enough to accommodate a gene expression cassette.

E1 deletions can be obtained starting at about base pair 342 going up to about base pair 3523 of Ad5, or a corresponding region from other adenoviruses.

Preferably, the deleted region involves removing a region from about base pair 450 to about base pair 3511 of Ad5, or a corresponding region from other adenoviruses. Larger E1 region deletions starting at about base pair 341 removes elements that facilitate virus packaging.

E3 deletions can be obtained starting at about base pair 27865 to about base pair 30995 of Ad5, or the corresponding region of other adenovectors. Preferably the deletion region involves removing a region from about base pair 28134 up to about base pair 30817 of Ad5, or the corresponding region of other adenovectors.

5

10

15

20

25

30

35

The combination of deletions to the E1 region and optionally the E3 region should be sufficiently large so that the overall size of the recombinant genome containing the gene expression cassette does not exceed about 105% of the wild type adenovirus genome. For example, as recombinant adenovirus Ad5 genomes increase size above about 105% the genome becomes unstable. (Bett et al., Journal of Virology 67:5911-5921, 1993.)

Preferably, the size of the recombinant adenovirus genome containing the gene expression cassette is about 85% to about 105% the size of the wild type adenovirus genome. In different embodiments, the size of the recombinant adenovirus genome containing the expression cassette is about 100% to about 105.2%, or about 100%, the size of the wild type genome.

Approximately 7,500 kb can be inserted into an adenovirus genome with a E1 and E3 deletion. Without any deletion, the Ad5 genome is 35,935 base pairs and the Ad6 genome is 35,759 base pairs.

Replication of first generation adenovectors can be performed by supplying the E1 gene products in *trans*. The E1 gene product can be supplied in *trans*, for example, by using cell lines that have been transformed with the adenovirus E1 region. Examples of cells and cells lines transformed with the adenovirus E1 region are HEK 293 cells, 911 cells, PERC.6TM cells, and transfected primary human aminocytes cells. (Graham *et al.*, *Journal of Virology 36*:59-72, 1977, Schiedner *et al.*, *Human Gene Therapy 11*:2105-2116, 2000, Fallaux *et al.*, *Human Gene Therapy 9*:1909-1917, 1998, Bout *et al.*, U.S. Patent No. 6,033,908.)

A Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette should be inserted into a recombinant adenovirus genome in the region corresponding to the deleted E1 region or the deleted E3 region. The expression cassette can have a parallel or anti-parallel orientation. In a parallel orientation the transcription direction

of the inserted gene is the same direction as the deleted E1 or E3 gene. In an antiparallel orientation transcription the opposite strand serves as a template and the transcription direction is in the opposite direction.

In an embodiment of the present invention the adenovector has a gene expression cassette inserted in the E1 deleted region. The vector contains:

5

10

15

20

25

30

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
- c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6 joined to the fourth region.

In another embodiment of the present invention the adenovector has an expression cassette inserted in the E3 deleted region. The vector contains:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the first region;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- d) a gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;

e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and

f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region.

In preferred different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

B. DNA Plasmid Vectors

5

10

15

20

25

30

DNA vaccine plasmid vectors contain a gene expression cassette along with elements facilitating replication and preferably vector selection. Preferred elements provide for replication in non-mammalian cells and a selectable marker. The vectors should not contain elements providing for replication in human cells or for integration into human nucleic acid.

The selectable marker facilitates selection of nucleic acids containing the marker. Preferred selectable markers are those that confer antibiotic resistance. Examples of antibiotic selection genes include nucleic acid encoding resistance to ampicillin, neomycin, and kanamycin.

Suitable DNA vaccine vectors can be produced starting with a plasmid containing a bacterial origin of replication and a selectable marker. Examples of bacterial origins of replication providing for higher yields include the ColE1 plasmid-derived bacterial origin of replication. (Donnelly et al., Annu. Rev. Immunol. 15:617-648, 1997.)

The presence of the bacterial origin of replication and selectable marker allows for the production of the DNA vector in a bacterial strain such as *E. coli*. The selectable marker is used to eliminate bacteria not containing the DNA vector.

III. AD6 RECOMBINANT NUCLEIC ACID

Ad6 recombinant nucleic acid comprises an Ad6 region substantially similar to an Ad6 region found in SEQ. ID. NO. 8, and a region not present in Ad6 nucleic acid. Recombinant nucleic acid comprising Ad6 regions have different uses such as in producing different Ad6 regions, as intermediates in the production of Ad6 based vectors, and as a vector for delivering a recombinant gene.

5

10

15

20

25

30

As depicted in Figure 9, the genomic organization of Ad6 is very similar to the genomic organization of Ad5. The homology between Ad5 and Ad6 is approximately 98%.

In different embodiments, the Ad6 recombinant nucleic acid comprises a nucleotide region substantially similar to E1A, E1B, E2B, E2A, E3, E4, L1, L2, L3, or L4, or any combination thereof. A substantially similar nucleic acid region to an Ad6 region has a nucleotide sequence identity of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or a nucleotide difference of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides. Techniques and embodiments for determining substantially similar nucleic acid sequences are described in Section I.B. supra.

Preferably, the recombinant Ad6 nucleic acid contains an expression cassette coding for a polypeptide not found in Ad6. Examples of expression cassettes include those coding for HCV regions and those coding for other types of polypeptides.

Different types of adenoviral vectors can be produced incorporating different amounts of Ad6, such as first and second generation adenovectors. As noted in Section II.A. *supra*. first generation adenovectors are defective in E1 and can replicate when E1 is supplied *in trans*.

Second generation adenovectors contain less adenoviral genome than first generation vectors and can be used in conjugation with complementing cell lines and/or helper vectors supplying adenoviral proteins. Second generation adenovectors are described in different references such as Russell, *Journal of General Virology* 81:2573-2604, 2000; Hitt et al., 1997, Human Ad vectors for Gene Transfer, Advances in Pharmacology, Vol 40 Academic Press.

In an embodiment of the present invention, the Ad6 recombinant nucleic acid is an adenovirus vector defective in E1 that is able to replicate when E1 is

supplied in trans. Expression cassettes can be inserted into a deleted E1 region and/or a deleted E3 region.

An example of an Ad6 based adenoviral vector with an expression cassette provided in a deleted E1 region comprises or consists of:

5

10

15

20

25

30

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
- c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- e) an optionally present fourth region from about base pair 28134 to about base pair 30817 corresponding to Ad5, or from about base pair 28157 to about base pair 30788 corresponding to Ad6, joined to the third region;
- f) a fifth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, wherein the fifth region is joined to the fourth region if the fourth region is present, or the fifth is joined to the third region if the fourth region is not present; and
- g) a sixth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fifth region;

wherein at least one Ad6 region is present.

In different embodiments of the invention, all of the regions are from Ad6; all of the regions expect for the first and second are from Ad6; and 1, 2, 3, or 4 regions selected from the second, third, fourth, and fifth regions are from Ad6.

An example of an Ad6 based adenoviral vector with an expression cassette provided in a deleted E3 region comprises or consists of:

a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the first region;

- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
 - d) a gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;
 - e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region;

wherein at least one Ad6 region is present.

5

10

15

20

25

30

In different embodiment of the invention, all of the regions are from Ad6; all of the regions expect for the first and second are from Ad6; and 1, 2, 3, or 4 regions selected from the second, third, fourth and fifth regions are from Ad6.

IV. VECTOR PRODUCTION

Vectors can be produced using recombinant nucleic acid techniques such as those involving the use of restriction enzymes, nucleic acid ligation, and homologous recombination. Recombinant nucleic acid techniques are well known in the art. (Ausubel, Current Protocols in Molecular Biology, John Wiley, 1987-1998, and Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, 1989.)

Intermediate vectors are used to derive a therapeutic vector or to transfer an expression cassette or portion thereof from one vector to another vector. Examples of intermediate vectors include adenovirus genome plasmids and shuttle vectors.

Useful elements in an intermediate vector include an origin of replication, a selectable marker, homologous recombination regions, and convenient restriction sites. Convenient restriction sites can be used to facilitate cloning or release of a nucleic acid sequence.

Homologous recombination regions provide nucleic acid sequence regions that are homologous to a target region in another nucleic acid molecule. The homologous regions flank the nucleic acid sequence that is being inserted into the target region. In different embodiments homologous regions are preferably about 150 to 600 nucleotides in length, or about 100 to 500 nucleotides in length.

An embodiment of the present invention describes a shuttle vector containing a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette, a selectable marker, a bacterial origin of replication, a first adenovirus homology region and a second adenovirus homologous region that target the expression cassette to insert in or replace an E1 region. The first and second homology regions flank the expression cassette. The first homology region contains at least about 100 base pairs substantially homologous to at least the right end (3' end) of a wild-type adenovirus region from about base pairs 4-450. The second homology contains at least about 100 base pairs substantially homologous to at least the left end (5' end) of Ad5 from about base pairs 3511-5792, or the corresponding region from another adenovirus.

Reference to "substantially homologous" indicates a sufficient degree of homology to specifically recombine with a target region. In different embodiments substantially homologous refers to at least 85%, at least 95%, or 100% sequence identity. Sequence identity can be calculated as described in Section I.B. supra.

One method of producing adenovectors is through the creation of an adenovirus genome plasmid containing an expression cassette. The pre-Adenovirus plasmid contains all the adenovirus sequences needed for replication in the desired complimenting cell line. The pre-Adenovirus plasmid is then digested with a restriction enzyme to release the viral ITR's and transfected into the complementing cell line for virus rescue. The ITR's must be released from plasmid sequences to allow replication to occur. Adenovector rescue results in the production on an adenovector containing the expression cassette.

A. Adenovirus Genome Plasmids

5

10

15

20

25

30

35

Adenovirus genome plasmids contain an adenovector sequence inside a longer-length plasmid (which may be a cosmid). The longer-length plasmid may contain additional elements such as those facilitating growth and selection in eukaryotic or bacterial cells depending upon the procedures employed to produce and maintain the plasmid. Techniques for producing adenovirus genome plasmids include those involving the use of shuttle vectors and homologous recombination, and those

involving the insertion of a gene expression cassette into an adenovirus cosmid. (Hitt et al., Methods in Molecular Genetics 7:13-30, 1995, Danthinne et al., Gene Therapy 7:1707-1714, 2000.)

Adenovirus genome plasmids preferably have a gene expression cassette inserted into a E1 or E3 deleted region. In an embodiment of the present invention, the adenovirus genome plasmid contains a gene expression cassette inserted in the E1 deleted region, an origin of replication, a selectable marker, and the recombinant adenovirus region is made up of:

5

15

20

25

30

- a) a first adenovirus region from about base pair 1 to about base
 450 corresponding to either Ad5 or Ad6;
 - b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
 - c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
 - e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region;
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region, and
 - g) an optionally present E3 region corresponding to all or part of the E3 region present in Ad5 or Ad6, which may be present for smaller inserts taking into account the overall size of the desired adenovector.

In another embodiment of the present invention the recombinant adenovirus genome plasmid has the gene expression cassette inserted in the E3 deleted region. The vector contains an origin of replication, a selectable marker, and the following:

a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;

c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;

5

10

15

20

25

30

35

- d) the gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region.

In different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

An embodiment of the present invention describes a method of making an adenovector involving a homologous recombination step to produce a adenovirus genome plasmid and an adenovirus rescue step. The homologous recombination step involves the use of a shuttle vector containing a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette flanked by adenovirus homology regions. The adenovirus homology regions target the expression cassette into either the E1 or E3 deleted region.

In an embodiment of the present invention concerning the production of an adenovirus genome plasmid, the gene expression cassette is inserted into a vector comprising: a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6; a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the second region; a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6,

joined to the second region; a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region; and a fifth adenovirus region from about 33967 to about 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region. The adenovirus genome plasmid should contain an origin of replication and a selectable marker, and may contain all or part of the Ad5 or Ad6 E3 region.

In different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

15 B. Adenovector Rescue

5

10

20

25

30

35

An adenovector can be rescued from a recombinant adenovirus genome plasmid using techniques known in the art or described herein. Examples of techniques for adenovirus rescue well known in the art are provided by Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, and Danthinne *et al.*, *Gene Therapy* 7:1707-1714, 2000.

A preferred method of rescuing an adenovector described herein involves boosting adenoviral replication. Boosting adenoviral replication can be performed, for example, by supplying adenoviral functions such as E2 proteins (polymerase, pre-terminal protein and DNA binding protein) as well as E4 orf6 on a separate plasmid. Example 10 *infra*. illustrates the boosting of adenoviral replication to rescue an adenovector containing a codon optimized Met-NS3-NS4A-NS4B-NS5B expression cassette.

V. PARTIAL-OPITIMIZED HCV ENCODING SEQUENCES

Partial optimization of HCV polyprotein encoding nucleic acid provides for a lesser amount of codons optimized for expression in a human than complete optimization. The overall objective is to provide the benefits of increased expression due to codon optimization, while facilitating the production of an adenovector containing HCV polyprotein encoding nucleic acid having optimized codons.

Complete optimization of an HCV polyprotein encoding sequence provides the most frequently observed human codon for each amino acid. Complete optimization can be performed using codon frequency tables well known in the art and using programs such as the BACKTRANSLATE program (Wisconsin Package version 10, Genetics Computer Group, GCG, Madison, Wisc.).

Partial optimization can be preformed on an entire HCV polyprotein encoding sequence that is present (e.g., NS3-NS5B), or one or more local regions that are present. In different embodiments the GC content for the entire HCV encoded polyprotein that is present is no greater than at least about 65%; and the GC content for one or more local regions is no greater than about 70%.

Local regions are regions present in HCV encoding nucleic acid, and can vary in size. For example, local regions can be about 60, about 70, about 80, about 90 or about 100 nucleotides in length.

10

15

20

25

30

35

Partial optimization can be achieved by initially constructing an HCV encoding polyprotein sequence to be partially optimized based on a naturally ocurring sequence. Alternatively, an optimized HCV encoding sequence can be used as basis of comparison to produce a partial optimized sequence.

VI. HCV COMBINATION TREATMENT

The HCV Met-NS3-NS4A-NS4B-NS5A-NS5B vaccine can be used by itself to treat a patient, can be used in conjunction with other HCV therapeutics, and can be used with agents targeting other types of diseases. Additional therapeutics include additional therapeutic agents to treat HCV and diseases having a high prevalence in HCV infected persons. Agents targeting other types of disease include vaccines directed against HIV and HBV.

Additional therapeutics for treating HCV include vaccines and non-vaccine agents. (Zein, Expert Opin. Investig. Drugs 10:1457-1469, 2001.) Examples of additional HCV vaccines include vaccines designed to elicit an immune response against an HCV core antigen and the HCV E1, E2 or p7 region. Vaccine components can be naturally occurring HCV polypeptides, HCV mimotope polypeptides or nucleic acid encoding such polypeptides.

HCV mimotope polypeptides contain HCV epitopes, but have a different sequence than a naturally occurring HCV antigen. A HCV mimotope can be fused to a naturally occurring HCV antigen. References describing techniques for producing mimotopes in general and describing different HCV mimotopes are

provided in Felici et al. U.S. Patent No. 5,994,083 and Nicosia et al., International Application Number WO 99/60132.

VII. PHARMACEUTICAL ADMINISTRATION

HCV vaccines can be formulated and administered to a patient using the guidance provided herein along with techniques well known in the art. Guidelines for pharmaceutical administration in general are provided in, for example, *Modern Vaccinology*, Ed. Kurstak, Plenum Med. Co. 1994; *Remington's Pharmaceutical Sciences 18th Edition*, Ed. Gennaro, Mack Publishing, 1990; and *Modern Pharmaceutics 2nd Edition*, Eds. Banker and Rhodes, Marcel Dekker, Inc., 1990, each of which are hereby incorporated by reference herein.

HCV vaccines can be administered by different routes such intravenous, intraperitoneal, subcutaneous, intramuscular, intradermal, impression through the skin, or nasal. A preferred route is intramuscular.

Intramuscular administration can be preformed using different techniques such as by injection with or without one or more electric pulses. Electric mediated transfer can assist genetic immunization by stimulating both humoral and cellular immune responses.

Vaccine injection can be performed using different techniques, such as by employing a needle or a needless injection system. An example of a needless injection system is a jet injection device. (Donnelly *et al.*, International Publication Number WO 99/52463.)

A. Electrically Mediated Transfer

5

10

15

20

35

Electrically mediated transfer or Gene Electro-Transfer (GET) can be performed by delivering suitable electric pulses after nucleic acid injection. (See Mathiesen, International Publication Number WO 98/43702). Plasmid injection and electroporation can be performed using stainless needles. Needles can be used in couples, triplets or more complex patterns. In one configuration the needles are soldered on a printed circuit board that is a mechanical support and connects the needles to the electrical field generator by means of suitable cables.

The electrical stimulus is given in the form of electrical pulses. Pulses can be of different forms (square, sinusoidal, triangular, exponential decay) and different polarity (monopolar of positive or negative polarity, bipolar). Pulses can be delivered either at constant voltage or constant current modality.

Different patterns of electric treatment can be used to introduce nucleic acid vaccines including HCV and other nucleic acid vaccines into a patient. Possible patterns of electric treatment include the following:

Treatment 1: 10 trains of 1000 square bipolar pulses delivered every other second, pulse length 0.2 msec/phase, frequency 1000 Hz, constant voltage mode, 45 Volts/phase, floating current.

Treatment 2: 2 trains of 100 square bipolar pulses delivered every other second, pulse length 2 msec/phase, frequency 100 Hz, constant current mode, 100 mA/phase, floating voltage.

Treatment 3: 2 trains of bipolar pulses at a pulse length of about 2 msec/phase, for a total length of about 3 seconds, where the actual current going through the tissue is fixed at about 50 mA.

Electric pulses are delivered through an electric field generator. A suitable generator can be composed of three independent hardware elements assembled in a common chassis and driven by a portable PC which runs the driving program. The software manages both basic and accessory functions. The elements of the device are: (1) signal generator driven by a microprocessor, (2) power amplifier and (3) digital oscilloscope.

The signal generator delivers signals having arbitrary frequency and shape in a given range under software control. The same software has an interactive editor for the waveform to be delivered. The generator features a digitally controlled current limiting device (a safety feature to control the maximal current output). The power amplifier can amplify the signal generated up to +/- 150 V. The oscilloscope is digital and is able to sample both the voltage and the current being delivered by the amplifier.

B. Pharmaceutical Carriers

5

10

15

20

25

30

35

Pharmaceutically acceptable carriers facilitate storage and administration of a vaccine to a subject. Examples of pharmaceutically acceptable carriers are described herein. Additional pharmaceutical acceptable carriers are well known in the art.

Pharmaceutically acceptable carriers may contain different components such a buffer, normal saline or phosphate buffered saline, sucrose, salts and polysorbate. An example of a pharmaceutically acceptable carrier is follows: 2.5-10 mM TRIS buffer, preferably about 5 mM TRIS buffer; 25-100 mM NaCl, preferably

about 75 mM NaCl; 2.5-10% sucrose, preferably about 5% sucrose; 0.01 -2 mM MgCl₂; and 0.001%-0.01% polysorbate 80 (plant derived). The pH is preferably from about 7.0-9.0, more preferably about 8.0. A specific example of a carrier contains 5 mM TRIS, 75 mM NaCl, 5% sucrose, 1 mM MgCl₂, 0.005% polysorbate 80 at pH 8.0.

C. Dosing Regimes

5

10

15

20

25

30

35

Suitable dosing regimens can be determined taking into account the efficacy of a particular vaccine and factors such as age, weight, sex and medical condition of a patient; the route of administration; the desired effect; and the number of doses. The efficacy of a particular vaccine depends on different factors such as the ability of a particular vaccine to produce polypeptide that is expressed and processed in a cell and presented in the context of MHC class I and II complexes.

HCV encoding nucleic acid administered to a patient can be part of different types of vectors including viral vectors such as adenovector, and DNA plasmid vaccines. In different embodiments concerning administration of a DNA plasmid, about 0.1 to 10 mg of plasmid is administered to a patient, and about 1 to 5 mg of plasmid is administered to a patient. In different embodiments concerning administration of a viral vector, preferably an adenoviral vector, about 105 to 1011 viral particles are administered to a patient, and about 107 to 1010 viral particles are administered to a patient.

Viral vector vaccines and DNA plasmid vaccines may be administered alone, or may be part of a prime and boost administration regimen. A mixed modality priming and booster inoculation involves either priming with a DNA vaccine and boosting with viral vector vaccine, or priming with a viral vector vaccine and boosting with a DNA vaccine.

Multiple priming, for example, about to 2-4 or more may be used. The length of time between priming and boost may typically vary from about four months to a year, but other time frames may be used. The use of a priming regimen with a DNA vaccine may be preferred in situations where a person has a pre-existing anti-adenovirus immune response.

In an embodiment of the present invention, $1x10^7$ to $1x10^{12}$ particles and preferably about $1x10^{10}$ to $1x10^{11}$ particles of adenovector is administered directly into muscle tissue. Following initial vaccination a boost is performed with an adenovector or DNA vaccine.

In another embodiment of the present invention initial vaccination is performed with a DNA vaccine directly into muscle tissue. Following initial vaccination a boost is performed with an adenovector or DNA vaccine.

Agents such as interleukin-12, GM-CSF, B7-1, B7-2, IP10, Mig-1 can be coadministered to boost the immune response. The agents can be coadministered as proteins or through use of nucleic acid vectors.

D. Heterologous Prime-Boost

5

10

15

20

25

30

Heterologous prime-boost is a mixed modality involving the use of one type of viral vector for priming and another type of viral vector for boosting. The heterologous prime-boost can involve related vectors such as vectors based on different adenovirus serotypes and more distantly related viruses such adenovirus and poxvirus. The use of poxvirus and adenovirus vectors to protect mice against malaria is illustrated by Gilbert *et al.*, *Vaccine* 20:1039-1045, 2002.

Different embodiments concerning priming and boosting involve the following types of vectors expressing desired antigens such as Met-NS3-NS4A-NS4B-NS5A-NS5B: Ad5 vector followed by Ad6 vector; Ad6 vector followed by Ad5 vector; Ad5 vector followed by poxvirus vector; poxvirus vector followed by Ad5 vector; Ad6 vector followed by poxvirus vector; and poxvirus vector followed by Ad6 vector.

The length of time between priming and boosting typically varies from about four months to a year, but other time frames may be used. The minimum time frame should be sufficient to allow for an immunological rest. In an embodiment, this rest is for a period of at least 6 months. Priming may involve multiple priming with one type of vector, such as 2-4 primings.

Expression cassettes present in a poxvirus vector should contain a promoter either native to, or derived from, the poxvirus of interest or another poxvirus member. Different strategies for constructing and employing different types of poxvirus based vectors including those based on vaccinia virus, modified vaccinia virus, avipoxvirus, raccoon poxvirus, modified vaccinia virus Ankara, canarypoxviruses (such as ALVAC), fowlpoxviruses, cowpoxviruses, and NYVAC are well known in the art. (Moss, Current Topics in Microbiology and Immunology 158:25-38, 1982; Earl et al., In Current Protocols in Molecular Biology, Ausubel et al. eds., New York: Greene Publishing Associates & Wiley Interscience;

35 1991:16.16.1-16.16.7, Child et al., Virology 174(2):625-9, 1990; Tartaglia et al.,

Virology 188:217-232, 1992; U.S. Patent Nos., 4,603,112, 4,722,848, 4,769,330, 5,110,587, 5,174,993, 5,185,146, 5,266,313, 5,505,941, 5,863,542, and 5,942,235.

E. Adjuvants

5

HCV vaccines can be formulated with an adjuvant. Adjuvants are particularly useful for DNA plasmid vaccines. Examples of adjuvants are alum, AlPO4, alhydrogel, Lipid-A and derivatives or variants thereof, Freund's incomplete adjuvant, neutral liposomes, liposomes containing the vaccine and cytokines, non-ionic block copolymers, and chemokines.

10

Non-ionic block polymers containing polyoxyethylene (POE) and polyxylpropylene (POP), such as POE-POP-POE block copolymers may be used as an adjuvant. (Newman et al., Critical Reviews in Therapeutic Drug Carrier Systems 15:89-142, 1998.) The immune response of a nucleic acid can be enhanced using a non-ionic block copolymer combined with an anionic surfactant.

15

20

25

A specific example of an adjuvant formulation is one containing CRL-1005 (CytRx Research Laboratories), DNA, and benzylalkonium chloride (BAK). The formulation can be prepared by adding pure polymer to a cold (<5°C) solution of plasmid DNA in PBS using a positive displacement pipette. The solution is then vortexed to solubilize the polymer. After complete solubilization of the polymer a clear solution is obtained at temperatures below the cloud point of the polymer (~6-7°C). Approximately 4 mM BAK is then added to the DNA/CRL-1005 solution in PBS, by slow addition of a dilute solution of BAK dissolved in PBS. The initial DNA concentration is approximately 6 mg/mL before the addition of polymer and BAK, and the final DNA concentration is about 5 mg/mL. After BAK addition the formulation is vortexed extensively, while the temperature is allowed to increase from ~ 2°C to above the cloud point. The formulation is then placed on ice to decrease the temperature below the cloud point. Then, the formulation is vortexed while the temperature is allowed to increase from ~2°C to above the cloud point. Cooling and mixing while the temperature is allowed to increase from ~2°C to above the cloud point is repeated several times, until the particle size of the formulation is about 200-500 nm, as measured by dynamic light scattering. The formulation is then stored on ice until the solution is clear, then placed in storage at -70°C. Before use, the formulation is allowed to thaw at room temperature.

F. Vaccine Storage

5

10

15

20

25

30

35

Adenovector and DNA vaccines can be stored using different types of buffers. For example, buffer A105 described in Example 9 *infra*. can be used to for vector storage.

Storage of DNA can be enhanced by removal or chelation of trace metal ions. Reagents such as succinic or malic acid, and chelators can be used to enhance DNA vaccine stability. Examples of chelators include multiple phosphate ligands and EDTA. The inclusion of non-reducing free radical scavengers, such as ethanol or glycerol, can also be useful to prevent damage of DNA plasmid from free radical production. Furthermore, the buffer type, pH, salt concentration, light exposure, as well as the type of sterilization process used to prepare the vials, may be controlled in the formulation to optimize the stability of the DNA vaccine.

VII. EXAMPLES

Examples are provided below to further illustrate different features of the present invention. The examples also illustrate useful methodology for practicing the invention. These examples do not limit the claimed invention.

Example 1: Met-NS3-NS4A-NS4B-NS5A-NS5B Expression Cassettes

Different gene expression cassettes encoding HCV NS3-NS4A-NS4B-NS5A-NS5B were constructed based on a 1b subtype HCV BK strain. The encoded sequences had either (1) an active NS5B sequence ("NS"), (2) an inactive NS5B sequence ("NSmut"), (3) a codon optimized sequence with an inactive NS5B sequence ("NSOPTmut"). The expression cassettes also contained a CMV promoter/enhancer and the BGH polyadenylation signal.

The NS nucleotide sequence (SEQ. ID. NO. 5) differs from HCV BK strain GenBank accession number M58335 by 30 out of 5952 nucleotides. The NS amino acid sequence (SEQ. ID. NO. 6) differs from the corresponding 1b genotype HCV BK strain by 7 out of 1984 amino acids. To allow for initiation of translation an ATG codon is present at the 5' end of the NS sequence. A TGA termination sequence is present at the 3' end of the NS sequence.

The NSmut nucleotide sequence (SEQ. ID. NO. 2, Figure 2), is similar to the NS sequence. The differences between NSmut and NS include NSmut having an altered NS5B catalytic site; an optimal ribosome binding site at the 5' end; and a TAAA termination sequence at the 3' end. The alterations in NS5B comprise bases

5138 to 5146, which encode amino acids 1711 to 1713. The alterations result in a change of amino acids GlyAspAsp into AlaAlaGly and creates an inactive form of the NS5B RNA-dependent RNA-polymerase NS5B.

The NSOPTmut sequence (SEQ. ID. NO. 3, Figure 3) was designed based on the amino acid sequence encoded by NSmut. The NSmut amino acid sequence was back translated into a nucleotide sequence with the GCG (Wisconsin Package version 10, Genetics Computer Group, GCG, Madison, Wisc.)

BACKTRANSLATE program. To generate a NSOPTmut nucleotide sequence where each amino acid is coded for by the corresponding most frequently observed human codon, the program was run choosing as parameter the generation of the most probable nucleotide sequence and specifying the codon frequency table of highly expressed human genes (human_high.cod) available within the GCG Package as translation scheme.

Example 2: Generation pV1Jns plasmid with NS, NSmut or NSOPTmut Sequences pV1Jns plasmids containing either the NS sequence, NSmut sequence or NSOPTmut sequences were generated and characterised as follows:

pV1Ins Plasmid with the NS Sequence

20

25

30

The coding region Met-NS3-NS4A-NS4B-NS5A and the coding region Met-NS3-NS4A-NS4B-NS5A-NS5B from a HCV BK type strain (Tomei *et al., J. Virol. 67*:4017-4026, 1993) were cloned into pcDNA3 plasmid (Invitrogen), generating pcD3-5a and pcD3-5b vectors, respectively. PcD3-5A was digested with Hind III, blunt-ended with Klenow fill-in and subsequently digested with Xba I, to generate a fragment corresponding to the coding region of Met-NS3-NS4A-NS4B-NS5A. The fragment was cloned into pV1Jns-poly, digested with Bgl II blunt-ended with Klenow fill-in and subsequently digested with Xba I, generating pV1JnsNS3-5A.

pV1Jns-poly is a derivative of pV1JnsA plasmid (Montgomery et al., DNA and Cell Biol. 12:777-783, 1993), modified by insertion of a polylinker containing recognition sites for XbaI, PmeI, PacI into the unique BglII and NotI restriction sites. The pV1Jns plasmid with the NS sequence (pV1JnsNS3-5B) was obtained by homologous recombination into the bacterial strain BJ5183, cotransforming pV1JNS3-5A linearized with XbaI and NotI digestion and a PCR fragment containing approximately 200 bp of NS5A, NS5B coding sequence and

approximately 60 bp of the BGH polyadenylation signal. The resulting plasmid represents pV1Jns-NS.

pV1Jns-NS can be summarized as follows:

Bases 1 to 1881 of pVIJnsA

5 an additional AGCTT

then the Met-NS3-NS5B sequence (SEQ. ID. NO. 5)

then the wt TGA stop

an additional TCTAGAGCGTTTAAACCCTTAATTAAGG (SEQ. ID.

NO. 14)

15

35

10 Bases 1912 to 4909 of pV1JnsA

pVIIns Plasmid with the NSmut Sequence

The V1JnsNS3-5A plasmid was modified at the 5' of the NS3 coding sequence by addition of a full Kozak sequence. The plasmid (V1JNS3-5Akozak) was obtained by homologous recombination into the bacterial strain BJ5183, cotransforming V1JNS3-5A linearized by AfIII digestion and a PCR fragment containing the proximal part of Intron A, the restriction site BglII, a full Kozak translation initiation sequence and part of the NS3 coding sequence.

The resulting plasmid (V1JNS3-5Akozak) was linearized with Xba I

digestion and co-transformed into the bacterial strain BJ5183 with a PCR fragment,
containing approximately 200 bp of NS5A, the NS5B mutated sequence, the strong
translation termination TAAA and approximately 60 bp of the BGH polyadenylation
signal. The PCR fragment was obtained by assembling two 22bp-overlapping
fragments where mutations were introduced by the oligonucleotides used for their
amplification. The resulting plasmid represents pV1Jns-NSmut.

pV1Jns-NSmut can be summarized as follows:

Bases 1 to 1882 of pV1JnsA

then the kozak Met-NS3-NS5B(mut) TAAA sequence (SEQ. ID. NO. 2)

an additional TCTAGA

30 Bases 1925 to 4909 of pV1JnsA

pVIJns Plasmid with the NSOPTmut Sequence

The human codon-optimized synthetic gene (NSOPTmut) with mutated NS5B to abrogate enzymatic activity, full Kozak translation initiation sequence and a strong translation termination was digested with BamHI and SalI

PCT/US02/32512 WO 03/031588

restriction sites present at the 5' and 3' end of the gene. The gene was then cloned into the BglII and SalI restriction sites present in the polylinker of pV1JnsA plasmid, generating pV1Jns-NSOPTmut.

pV1Jns-NSOPTmut can be summarized as follows:

5 Bases 1 to 1881 of pV1JnsA

an additional C

then

kozak Met-NS3-NS5B(optmut) TAAA sequence (SEQ. ID. NO. 3)

an additional TTTAAATGTTTAAAC (SEQ. ID. NO. 15)

Bases

1905 to 4909 of pV1JnsA

10

15

20

30

35

Plasmids Characterization

Expression of HCV NS proteins was tested by transfection of HEK 293 cells, grown in 10% FCS/DMEM supplemented by L-glutamine (final 4 mM). Twenty-four hours before transfection, cells were plated in 6-well 35 mm diameter, to reach 90-95% confluence on the day of transfection. Forty nanograms of plasmid DNA (previously assessed as a non-saturating DNA amount) were co-transfected with 100 ng of pRSV-Luc plasmid containing the luciferase reporter gene under the control of Rous sarcoma virus promoter, using the LIPOFECTAMINE 2000 reagent. Cells were kept in a CO₂ incubator for 48 hours at 37 °C.

Cell extracts were prepared in 1% Triton/TEN buffer. The extracts were normalized for Luciferase activity, and run in serial dilution on 10% SDSacrylamide gel. Proteins were transferred on nitrocellulose and assayed with antibodies directed against NS3, NS5A and NS5B to assess strength of expression and correct proteolytic cleavage. Mock-transfected cells were used as a negative control.

Results from representative experiments testing pV1JnsNS, pV1JnsNSmut and 25 pV1JnsNSOPTmut are shown in Figure 12.

Example 3: Mice Immunization with Plasmid DNA Vectors

The DNA plasmids pV1Jns-NS, pV1Jns-NSmut and pV1Jns-NSOPT mut were injected in different mice strains to evaluate their potential to elicit anti-HCV immune responses. Two different strains (Balb/C and C57Black6, N=9-10) were injected intramuscularly with 25 or 50 µg of DNA followed by electrical pluses. Each animal received two doses at three weeks interval.

Humoral immune response elicited in C57Black6 mice against the NS3 protein was measured in post dose two sera by ELISA on bacterially expressed NS3

protease domain. Antibodies specific for the tested antigen were detected in animals immunized with all three vectors with geometric mean titers (GMT) ranging from 94000 to 133000 (Tables 1-3).

5

Table 1: pV1ins-NS

						-				GMT
Mice n.	1	2	3	4	5	6	7	8	9	
Titer	105466	891980	78799	39496	543542	182139	32351	95028	67800	94553

Table 2: pV1jns-NSmut

10

								·····		•	GMT
Mice n.	11	12	13	14	15	16	17	18	19	20	-
Titer	202981	55670	130786	49748	17672	174958	44304	37337	78182	193695	75083

Table 3: pV1jns-NSOPTmut

			-				****			•	GMT
Mice n.	21	22	23	24	25	26	27	28	29	30	
Titer	310349	43645	63496	82174	630778	297259	66861	146735	173506	77732	133165

15

20

25

A T cell response was measured in C57Black6 mice immunized with two intramuscular injections at three weeks interval with 25 μ g of plasmid DNA. Quantitative ELIspot assay was performed to determine the number of IFN γ secreting T cells in response to five pools of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence. Specific CD8+ response was analyzed by the same assay using a 20mer peptide encompassing a CD8+ epitope for C57Black6 mice (pep1480).

Cells secreting IFN γ in an antigen specific-manner were detected using a standard ELIspot assay. T cell response in C57Black6 mice immunized with two intramuscular injections at three weeks interval with 50 μ g of plasmid DNA, was

analyzed by the same ELIspot assay measuring the number of IFN γ secreting T cells in response to five pools of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence.

Spleen cells were prepared from immunized mice and re-suspended in R10 medium (RPMI 1640 supplemented with 10% FCS, 2 mM L-Glutamine, 50 U/ml-50μg/ml Penicillin/Streptomycin, 10 mM Hepes, 50 μM 2-mercapto-ethanol). Multiscreen 96-well Filtration Plates (Millipore, Cat. No. MAIPS4510, Millipore Corporation, 80 Ashby Road Bedford, MA) were coated with purified rat anti-mouse INFγ antibody (PharMingen, Cat. No. 18181D, PharmiMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA). After overnight incubation, plates were washed with PBS 1X/0.005% Tween and blocked with 250 μl/well of R10 medium.

Splenocytes from immunized mice were prepared and incubated for twenty-four hours in the presence or absence of 10 μM peptide at a density of 2.5 X 10⁵/well or 5 X 10⁵/well. After extensive washing (PBS 1X/0.005% Tween), biotinylated rat anti-mouse IFNγ antibody (PharMingen, Cat. No. 18112D, PharMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA) was added and incubated overnight at 4° C. For development, streptavidin-AKP (PharMingen, Cat. No. 13043E, PharMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA) and 1-StepTM NBT-BCIP development solution (Pierce, Cat. No. 34042, Pierce, P.O. Box 117, Rockford, IL 61105 USA) were added.

Pools of 20mer overlapping peptides encompassing the entire sequence of the HCV BK strain NS3 to NS5B were used to reveal HCV-specific IFNγ-secreting T cells. Similarly a single 20mer peptide encompassing a CD8+ epitope for C57Black6 mice was used to detect CD8 response. Representative data from groups of C57Black6 and Balb/C mice (N=9-10) immunized with two injections of 25 or 50 μg of plasmid vectors pV1Jns-NS, pV1Jns-NSmut and pV1Jns-NSOPTmut are shown in Figures 13A and 13B.

30 Example 4: Immunization of Rhesus Macaques

15

20

25

Rhesus macaques (N=3) were immunized by intramuscular injection with 5mg of plasmid pV1Jns-NSOPTmut in 7.5mg/ml CRL1005, Benzalkonium chloride 0.6 mM. Each animal received two doses in the deltoid muscle at 0, and 4 weeks.

CMI was measured at different time points by IFN- γ ELISPOT. This assay measures HCV antigen-specific CD8+ and CD4+ T lymphocyte responses, and can be used for a variety of mammals, such as humans, rhesus monkeys, mice, and rats.

The use of a specific peptide or a pool of peptides can simplify antigen presentation in CTL cytotoxicity assays, interferon-gamma ELISPOT assays and interferon-gamma intracellular staining assays. Peptides based on the amino acid sequence of various HCV proteins (core, E2, NS3, NS4A, NS4B, NS5A, NS5B) were prepared for use in these assays to measure immune responses in HCV DNA and adenovirus vector vaccinated rhesus monkeys, as well as in HCV-infected humans. The individual peptides are overlapping 20-mers, offset by 10 amino acids. Large pools of peptides can be used to detect an overall response to HCV proteins while smaller pools and individual peptides may be used to define the epitope specificity of a response.

15

20

25

30

10

5

IFNy ELISPOT

The IFNγ-ELISPOT assay provides a quantitative determination of HCV-specific T lymphocyte responses. PBMC are serially diluted and placed in microplate wells coated with anti-rhesus IFN-γ antibody (MD-1 U-Cytech). They are cultured with a HCV peptide pool for 20 hours, resulting in the restimulation of the precursor cells and secretion of IFN-γ. The cells are washed away, leaving the secreted IFN bound to the antibody-coated wells in concentrated areas where the cells were sitting. The captured IFN is detected with biotinylated anti-rhesus IFN antibody (detector Ab U-Cytech) followed by alkaline phosphatase-conjugated streptavidin (Pharmingen 13043E). The addition of insoluble alkaline phosphatase substrate results in dark spots in the wells at the sites where the cells were located, leaving one spot for each T cell that secreted IFN-γ.

The number of spots per well is directly related to the precursor frequency of antigen-specific T cells. Gamma interferon was selected as the cytokine visualized in this assay (using species specific anti-gamma interferon monoclonal antibodies) because it is the most common, and one of the most abundant cytokines synthesized and secreted by activated T lymphocytes. For this assay, the number of spot forming cells (SFC) per million PBMCs is determined for samples in the

· ...

presence and absence (media control) of peptide antigens. Data from Rhesus macaques on PBMC from post dose two material are shown in Table 4.

Table 4

		PV1J-NSOPTmut	
Pep pools	21G	99C161	99C166
F (NS3p)	8	10	170
G (NS3h)	7	592	229
H (NS4)	3	14	16
I (NS5a)	5	71	36
L (NS5b)	14	23	11
M (NS5b)	3	35	8
DMSO	2	4	5

INFγELISPOT on PBMC from Rhesus monkeys immunized with two injections of 5 mg DNA/dose in OPTIVAX/BAK of plasmid pV1Jns-NSOPTmut. Data are expressed as SFC7 106 PBMC.

Example 5: Construction of Ad6 Pre-Adenovirus Plasmids

Ad6 pre-adenovirus plasmids were obtained as follows:

10

15

20

Construction of pAd6 E1-E3+ Pre-adenovirus Plasmid

An Ad6 based pre-adenovirus plasmid which can be used to generate first generation Ad6 vectors was constructed either taking advantage of the extensive sequence identity (approx. 98%) between Ad5 and Ad6 or containing only Ad6 regions. Homologous recombination was used to clone wtAd6 sequences into a bacterial plasmid.

A general strategy used to recover pAd6E1-E3+ as a bacterial plasmid containing Ad5 and Ad6 regions is illustrated in Figure 10. Cotransformation of BJ 5183 bacteria with purified wt Ad6 viral DNA and a second DNA fragment termed the Ad5 ITR cassette resulted in the circularization of the viral genome by homologous recombination. The ITR cassette contains sequences from the right (bp 33798 to 35935) and left (bp 1 to 341 and bp 3525 to 5767) end of the Ad5 genome separated by plasmid sequences containing a bacterial origin of replication and an ampicillin resistance gene. The ITR cassette contains a deletion of E1 sequences from

Ad5 342 to 3524. The Ad5 sequences in the ITR cassette provide regions of homology with the purified Ad6 viral DNA in which recombination can occur.

Potential clones were screened by restriction analysis and one clone was selected as pAd6E1-E3+. This clone was then sequenced in it entirety. pAd6E1-E3+ contains Ad5 sequences from bp 1 to 341 and from bp 3525 to 5548, Ad6 bp 5542 to 33784, and Ad5 bp 33967 to 35935 (bp numbers refer to the wt sequence for both Ad5 and Ad6). pAd6E1-E3+ contains the coding sequences for all Ad6 virion structural proteins which constitute its serotype specificity.

A general strategy used to recover pAd6E1-E3+ as a bacterial plasmid

containing Ad6 regions is illustrated in Figure 11. Cotransformation of BJ 5183

bacteria with purified wt Ad6 viral DNA and a second DNA fragment termed the Ad6

ITR cassette resulted in the circularization of the viral genome by homologous

recombination. The ITR cassette contains sequences from the right (bp 35460 to

35759) and left (bp 1 to 450 and bp 3508 to 3807) end of the Ad6 genome separated

by plasmid sequences containing a bacterial origin of replication and an ampicillin

resistance gene. These three segments were generated by PCR and cloned

sequentially into pNEB193, generating pNEBAd6-3 (the ITR cassette). The ITR

cassette contains a deletion of E1 sequences from Ad5 451 to 3507. The Ad6

sequences in the ITR cassette provide regions of homology with the purified Ad6 viral

DNA in which recombination can occur.

Construction of pAd6 E1-E3- pre-adenovirus plasmids

25

Ad6 based vectors containing A5 regions and deleted in the E3 region were constructed starting with pAd6E1-E3+ containing Ad5 regions. A 5322 bp subfragment of pAd6E1-E3+ containing the E3 region (Ad6 bp 25871 to 31192) was subcloned into pABS.3 generating pABSAd6E3. Three E3 deletions were then made in this plasmid generating three new plasmids pABSAd6E3(1.8Kb) (deleted for Ad6 bp 28602 to 30440), pABSAd6E3(2.3Kb) (deleted for Ad6 bp 28157 to 30437) and pABSAd6E3(2.6Kb) (deleted for Ad6 bp 28157 to 30788). Bacterial recombination was then used to substitute the three E3 deletions back into pAd6E1-E3+ generating the Ad6 genome plasmids pAd6E1-E3-1.8Kb, pAd6E1-E3-2.3Kb and pAd6E1-E3-2.6Kb.

Example 6: Generation of Ad5 Genome Plasmid with the NS Sequence

5

10

15

20

25

30

A pcDNA3 plasmid (Invitrogen) containing the coding region NS3-NS4A-NS4B-NS5A was digested with *Xmn*I and *Nru*I restriction sites and the DNA fragment containing the CMV promoter, the NS3-NS4A-NS4B-NS5A coding sequence and the Bovine Growth Hormone (BGH) polyadenylation signal was cloned into the unique *EcorV* restriction site of the shuttle vector pDelE1Spa, generating the Sva3-5A vector.

A pcDNA3 plasmid containing the coding region NS3-NS4A-NS4B-NS5A-NS5B was digested with *XmnI* and *EcorI* (partial digestion), and the DNA fragment containing part of NS5A, NS5B gene and the BGH polyadenylation signal was cloned into the Sva3-5A vector, digested *EcorI* and *BgIII* blunted with Klenow, generating the Sva3-5B vector.

The Sva3-5B vector was finally digested *SspI* and *Bst*1107I restriction sites and the DNA fragment containing the expression cassette (CMV promoter, NS3-NS4A-NS4B-NS5A-NS5B coding sequence and the BGH polyadenylation signal) flanked by adenovirus sequences was co-transformed with pAd5HVO (E1-,E3-) ClaI linearized genome plasmid into the bacterial strain BJ5183, to generate pAd5HVONS. pAd5HVO contains Ad5 bp 1 to 341, bp 3525 to 28133 and bp 30818 to 35935.

Example 7: Generation of Adenovirus Genome Plasmids with the NSmut Sequence
Adenovirus genome plasmids containing an NS-mut sequence were
generated in an Ad5 or Ad6 background. The Ad6 background contained Ad5 regions
at bases 1 to 450, 3511 to 5548 and 33967 to 35935.

pV1JNS3-5Akozak was digested with *Bgl*II and *Xba*I restriction enzymes and the DNA fragment containing the Kozak sequence and the sequence coding NS3-NS4A-NS4B-NS5A was cloned into a *Bgl*II and XbaI digested polypMRKpdelE1 shuttle vector. The resulting vector was designated shNS3-5Akozak.

PolypMRKpdelE1 is a derivative of RKpdelE1(Pac/pIX/pack450) + CMVmin+BGHpA(str.) modified by the insertion of a polylinker containing recognition sites for BglII, PmeI, SwaI, XbaI, SalI, into the unique BglII restriction site present downstream the CMV promoter. MRKpdelE1(Pac/pIX/pack450) + CMVmin + BGHpA(str.) contains Ad5 sequences from bp 1 to 5792 with a deletion of E1 sequences from bp 451 to 3510. The human CMV promoter and BGH polyadenylation signal were inserted into the E1 deletion in an E1 parallel orientation with a unique BglII site separating them.

The NS5B fragment, mutated to abrogate enzymatic activity and with a strong translation termination at the 3' end, was obtained by assembly PCR and inserted into the shNS3-5Akozak vector via homologous recombination, generating polypMRKpdelE1NSmut. In polypMRKpdelE1NSmut the NS-mut coding sequence is under the control of CMV promoter and the BGH polyadenylation signal is present downstream.

The gene expression cassette and the flanking regions which contain adenovirus sequences allowing homologous recombination were excised by digestion with *PacI* and *Bst*1107I restriction enzymes and co-transformed with either pAd5HVO (E1-,E3-) or pAd6E1-E3-2.6Kb *ClaI* linearized genome plasmids into the bacterial strain BJ5183, to generate pAd5HVONSmut and pAd6E1-,E3-NSmut, respectively.

10

15

pAd6E1-E3-2.6Kb contains Ad5 bp 1 to 341 and from bp 3525 to 5548, Ad6 bp 5542 to 28157 and from bp 30788 to 33784, and Ad5 bp 33967 to 35935 (bp numbers refer to the wt sequence for both Ad5 and Ad6). In both plasmids the viral ITR's are joined by plasmid sequences that contain the bacterial origin of replication and an ampicillin resistance gene.

Example 8: Generation of Adenovirus Genome Plasmids with the NSOPTmut

The human codon-optimized synthetic gene (NSOPTmut) provided by SEQ. ID. NO. 3 cloned into a pCRBlunt vector (Invitrogen) was digested with BamH1 and SalI restriction enzymes and cloned into BglII and SalI restriction sites present in the shuttle vector polypMRKpdelE1. The resulting clone (polypMRKpdelE1NSOPTmut) was digested with PacI and Bst1107I restriction enzymes and co-transformed with either pAd5HVO (E1-,E3-) or pAd6E1-E3-2.6Kb

ClaI linearized genome plasmids, into the bacterial strain BJ5183, to generate pAd5HVONSOPTmut and pAd6E1-,E3-NSOPTmut, respectively.

Example 9: Rescue and Amplification of Adenovirus Vectors

Adenovectors were rescued in Per.6 cells. Per.C6 were grown in 10% FCS / DMEM supplemented by L-glutamine (final 4mM), penicillin/streptomycin (final 100 IU/ml) and 10 mM MgCl₂. After infection, cells were kept in the same medium supplemented by 5% horse serum (HS). For viral rescue, 2.5 X 10⁶ Per.C6 were plated in 6 cm Ø Petri dishes.

Twenty-four hours after plating, cells were transfected by calcium phosphate method with 10 μ g of the *Pac I* linearized adenoviral DNA. The DNA precipitate was left on the cells for 4 hours. The medium was removed and 5% HS/DMEM was added.

Cells were kept in a CO₂ incubator until a cytopathic effect was visible (1 week). Cells and supernatant were recovered and subjected to 3X freeze/thawing cycles (liquid nitrogen / water bath at 37°C). The lysate was centrifuged at 3000 rpm at - 4°C for 20 minutes and the recovered supernatant (corresponding to a cell lysate containing virus passed on cells only once; P1) was used, in the amount of 1 ml/dish, to infect 80-90% confluent Per.C6 in 10 cm ø Petri dishes. The infected cells were incubated until a cytopathic effect was visible, cells and supernatant recovered and the lysate prepared as described above (P2).

P2 lysate (4 ml) were used to infect 2 X 15 cm ø Petri dishes. The lysate recovered from this infection (P3) was kept in aliquots at -80°C as a stock of virus to be used as starting point for big viral preparations. In this case, 1 ml of the stock was enough to infect 2 X 15 cm ø Petri dishes and resulting lysate (P4) was used for the infection of the Petri dishes devoted to the large scale infection.

Further amplification was obtained from the P4 lysate which was diluted in medium without FCS and used to infect 30 X 15 cm Ø Petri dishes (with Per.C6 80%-90% confluent) in the amount of 10 ml/dish. Cells were incubated 1 hour in the CO₂ incubator, mixing gently every 20 minutes. 12 ml / dish of 5% HS / DMEM was added and cells were incubated until a cytopathic effect was visible (about 48 hours).

Cells and supernatant were collected and centrifuged at 2K rpm for 20 minutes at 4° C. The pellet was resuspended in 15 ml of 0.1 M Tris pH=8.0. Cells were lysed by 3X freeze/thawing cycles (liquid nitrogen / water bath at 37° C). 150 μ l of 2 M MgCl₂ and 75 μ l of DNAse (10 mg of bovine pancreatic deoxyribonuclease I in 10 ml of 20 mM Tris-HCl pH= 7.4, 50 mM NaCl, 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, 50% glycerol) were added. After a 1 hour incubation at 37° C in a water bath (vortex every 15 minutes) the lysate was centrifuged at 4K rpm for 15 minutes at 4° C. The recovered supernatant was ready to be applied on CsCl gradient.

The CsCl gradients were prepared in SW40 ultra-clear tubes as

follows:

5

10

15

20

25

30

0.5 ml of 1.5d CsCl

35 3 ml of 1.35d CsCl

3 ml of 1.25d CsCl

5

10

25

35

5-ml/ tube of viral supernatant was applied.

If necessary, the tubes were topped up with 0.1 M tris-Cl pH=8.0. Tubes were centrifuged at 35K rpm for 1 hour at -10°C with rotor SW40. The viral bands (located at the 1.25/1.35 interface) were collected using a syringe.

The virus was transferred into a new SW40 ultraclear tube and 1.35d CsCl was added to top the tube up. After centrifugation at 35K rpm for 24 hours at 10° C in the rotor SW40, the virus was collected in the smallest possible volume and dialyzed extensively against buffer A105 (5 mM Tris, 5% sucrose, 75 mM NaCl, 1 mM MgCl₂, 0.005% polysorbate 80 pH=8.0). After dialysis, glycerol was added to final 10% and the virus was stored in aliquots at -80° C.

Example 10: Enhanced Adenovector Rescue

First generation Ad5 and Ad6 vectors carrying HCV NSOPTmut

transgene were found to be difficult to rescue. A possible block in the rescue process
might be attributed to an inefficient replication of plasmid DNA that is a sub-optimal
template for the replication machinery of adenovirus. The absence of the terminal
protein linked to the 5'ends of the DNA (normally present in the viral DNA),
associated with the very high G-C content of the transgene inserted in the E1 region of
the vector, may be causing a substantial reduction in replication rate of the plasmidderived adenovirus.

To set up a more efficient and reproducible procedure for rescuing Ad vectors, an expression vector (pE2; Figure 19) containing all E2 proteins (polymerase, pre-terminal protein and DNA binding protein) as well as E4 orf6 under the control of tet-inducible promoter was employed. The transfection of pE2 in combination with a normal preadeno plasmid in PerC6 and in 293 leads to a strong increase of Ad DNA replication and to a more efficient production of complete infectious adenovirus particles.

30 Plasmid Construction

pE2 is based on the cloning vector pBI (CLONTECH) with the addition of two elements to allow episomal replication and selection in cell culture: (1) the EBV-OriP (EBV [nt] 7421-8042) region permitting plasmid replication in synchrony with the cell cycle when EBNA-1 is expressed and (2) the hygromycin-B phosphotransferase (HPH)-resistance gene allowing a positive selection of

transformed cells. The two transcriptional units for the adenoviral genes E2 a and b and E4-Orf6 were constructed and assembled in pE2 as described below.

The Ad5-Polymerase Clal/SphI fragment and the Ad5-pTP Acc65/EcoRV fragment were obtained from pVac-Pol and pVac-pTP (Stunnemberg et al. NAR 16:2431-2444, 1988). Both fragments were filled with Klenow and cloned into the SalI (filled) and EcoRV sites of pBI, respectively obtaining pBI-Pol/pTP.

5

10

15

20

25

30

35

EBV-OriP element from pCEP4 (Invitrogen) was first inserted within two chicken β-globin insulator dimers by cloning it into *BamHI* site of pJC13-1 (Chung *et al.*, *Cell 74(3)*:505-14, 1993). HS4-OriP fragment from pJC13-OriP was then cloned inside pSA1mv (a plasmid containing tk-Hygro-B resistance gene expression cassette as well as Ad5 replication origin), the ITR's arranged as head-to-tail junction, obtained by PCR from pFG140 (Graham, *EMBO J. 3*:2917-2922, 1984) using the following primers: 5'-TCGAATCGATACGCGAACCTACGC-3' (SEQ. ID. NO. 16) and 5'-TCGACGTGTCGACTTCGAAGCGCACACCAAAAACGTC-3' (SEQ. ID. NO. 17), thus generating pMVHS4Orip. A DNA fragment from pMVHS4Orip, containing the insulated OriP, Ad5 ITR junction and tk-HygroB cassette, was then inserted into pBI-Pol/pTP vector restricted *Asel/AatII* generating pBI-Pol/pTPHS4.

To construct the second transcriptional unit expressing Ad5-Orf6 as well as Ad5-DBP, E4orf6 (Ad 5 [nt] 33193-34077) obtained by PCR was first inserted into pBI vector, generating pBI-Orf6. Subsequently, DBP coding DNA sequence (Ad 5 [nt] 22443-24032) was inserted into pBI-Orf6 obtaining the second bi-directional Tet-regulated expression vector (pBI-DBP/E4orf6). The original polyA signals present in pBI were substituted with BGH and SV40 polyA.

pBI-DBP/E4orf6 was then modified by inserting a DNA fragment containing the Adeno5-ITRs arranged in head-to-tail junction plus the hygromicin B resistance gene obtained from plasmid pSA-1mv. The new plasmid pBI-DBP/E4orf6shuttle was then used as donor plasmid to insert the second tet-regulated transcriptional unit into pBI-Pol/pTPHS4 by homologous recombination using *E. coli* strain BJ5183 obtaining pE2.

Cell lines, Transfections and Virus Amplification

PerC6 cells were cultured in Dulbecco's modified Eagle's Medium (DMEM) plus 10% fetal bovine serum (FBS), 10 mM MgCl₂, penicillin (100 U/ml), streptomycin (100 μ g/ml) and 2 mM glutamine.

All transient transfections were performed using Lipofectamine2000 (Invitrogen) as described by the manufacturer. 90% confluent PERC.6TM planted in 6-cm plates were transfected with 3.5 μg of Ad5/6NSOPTmut pre-adeno plasmids, digested with PacI, alone or in combination with 5 μg pE2 plus 1 μg pUHD52.1. pUHD52.1 is the expression vector for the reverse tet transactivator 2 (rtTA2) (Urlinger et al., Proc. Natl. Acad. Sci. U.S.A. 97(14):7963-7968, 2000). Upon transfection, cells were cultivated in the presence of 1 μg/ml of doxycycline to activate pE2 expression. 7 days post-transfection cells were harvested and cell lysate was obtained by three cycles of freeze-thaw. Two ml of cell lysate were used to infect a second 6-cm dish of PerC6. Infected cells were cultivated until a full CPE was observed then harvested. The virus was serially passaged five times as described above, then purified on CsCl gradient. The DNA structure of the purified virus was controlled by endonuclease digestion and agarose gel electrophoresis analysis and compared to the original pre-adeno plasmid restriction pattern.

15

20

25

30

35

10

5

Example 11: Partial Optimizeation of HCV Polyprotein Encoding Nucleic acid

Partial optimization of HCV polyprotein encoding nucleic acid was performed to facilitate the production of adenovectors containing codons optimized for expression in a human host. The overall objective was to provide for increased expression due to codon optimization, while facilitating the production of an adenovector encoding HCV polyprotein.

Several difficulties were encountered in producing an adenovector encoding HCV polyprotein with codons optimized for expression in a human host. An adenovector containing an optimized sequence (SEQ. ID. NO. 3) was found to be more difficult to synthesize and rescue than an adenovector containing a non-optimized sequence (SEQ. ID. NO. 2).

The difficulties in producing an adenovector containing SEQ. ID. NO. 3 were attributed to a high GC content. A particularly problemetic region was the region at about position 3900 of NSOPTmut (SEQ. ID. NO. 3).

Alternative versions of optimized HCV encoding nucleic acid sequence were designed to facilitate its use in an adenovector. The alternative versions, compared to NSOPTmut, were designed to have a lower overall GC content, to reduce/avoid the presence of potentially problematic motifis of consecutive G's or C's, while maintaining a high level of codon optimization to allow improved expression of the encoded polyprotein and the individual cleavage products.

A starting point for the generation of a suboptimally codon-optimized sequence is the coding region of the NSOPTmut nucleotide sequence (bases 7 to 5961 of SEQ. ID. NO. 3). Values for codon usage frequencies (normalized to a total of 1.0 for each amino acid) were taken from the file human_high.cod available in the Wisconsin Package Version 10.3 (Accelrys Inc., a wholly owned subsidiary of Pharmacopeia, Inc).

5

10

15

20

25

30

35

To reduce the local and overall GC content a table defining preferred codon substitutions for each amino acid was manually generated. For each amino acid the codon having 1) a lower GC content as compared to the most frequent codon and 2) a relativly high observed codon usage frequency (as defined in human_high.cod) was choosen as the replacement codon. For example for Arg the codon with the highest frequency is CGC. Out of the other five alternative codons encoding Arg (CGG, AGG, AGA, CGT, CGA) three (AGG, CGT, CGA) reduce the GC content by 1 base, one (AGA) by two bases and one (CGG) by 0 bases. Since the AGA codon is listed in human_high.cod as having a relatively low usage frequency (0.1), the codon substituting CGC was therefore choosen to be AGG with a relative frequency of 0.18. Similar criteria were applied in order to establish codon replacements for the other amino acids resulting in the list shown in Table 5. Parameters applied in the following optimization procedure were determined empirically such that the resulting sequence maintained a considerably improved codon usage (for each amino acid) and the GC content (overall and in form of local stretches of consecutive G's and/or C's) was decreased.

Two examples of partial optimized HCV encoding sequences are provided by SEQ. ID. NO. 10 and SEQ. ID. NO. 11. SEQ. ID. NO. 10 provides a HCV encoding sequence that is partially optimized throughout. SEQ. ID. NO. 11 provides an HCV encoding sequence fully optimized for codon usage with the exception of a region that was partially optimized.

Codon optimization was performed using the following procedure:

Step 1) The coding region of the input fully optimized NSOPTmut sequence was analyzed using a sliding window of 3 codons (9 bases) shifting the window by one codon after each cycle. Whenever a stretch containing 5 or more consecutive C's and/or G's was detected in the window the following replacement rule was applied: Let N indicate the number of codon replacements previously performed. If N is odd replace the middle codon in the window with the codon specified in Table 5, if N is even replace the third terminal codon in the window with the codon

specified in a codon optimization table such as human_high.cod. If Leu or Val is present at the second or third codon do not apply any replacement in order not to introduce Leu or Val codons with very low relative codon usage frequency (see, for example, human_high.cod). In the following cycle analysis of the shifted window was then applied to a sequence containing the replacements of the previous cycle.

5

10

15

20

25

30

The alternating replacement of the middle and terminal codon in the 3 codon window was found empirically to give a more satisfying overall maintenance of optimized codon usage while also reducing GC content (as judged from the final sequence after the procedure). In general, however, the precise replacement strategy depends on the amino acid sequence encoded by the nucelotide sequence under analysis and will have to be determined empirically.

Step 2) The sequence containing all the codon replacements performed during step 1) was then subjected to an additional analysis using a sliding window of 21 codons (63 bases) in length: according to an adjustable parameter the overall GC content in the window was determined. If the GC content in the window was higher than 70% the following codon replacement strategy was applied: In the window replace the codons for the amino acids Asn, Asp, Cys, Glu, His, Ile, Lys, Phe, Tyr by the codons given in Table 5. Restriction of the replacement to this set of amino acids was motivated by the fact that a) the replacement codon still has an accetably high frequency of usage in human_high.cod and b) the average overall human codon usage in CUTG for the replacement codon is nearly as high as the most frequent codon. In the following cycle analysis of the shifted window is then applied to a sequence containing the replacements of the previous cycle.

The threshold 70% was determined empirically by compromising between an overall reduction in GC content and maintenance of a high codon optimization for the individual amino acids. As in step 1) the precise replacement strategy (choice of amino acids and GC content threshold value) will again depend on the amino acid sequence encoded by the nucleotide sequence under analysis and will have to be determined empirically.

Step 3) The sequence generated by steps 1) and 2) was then manually edited and additional codons were changed according to the following criteria:

Regions still having a GC content higher than 70% over a window of 21 codons were examined manually and a few codons were replaced again following the scheme given in Table 5.

Subsequent steps were performed to provide for useful restriction sites, remove possible open reading frames on the complementary strand, to add homologous recombinant regions, to add a Kozac signal, and to add a terminator. These steps are numbered 4-7

Step 4) The sequence generated in step 3 was examined for the absence of certain restriction sites (BgIII, PmeI and XbaI) and presence of only 1 StuI site to allow a subsequent cloning strategy using a subset of restriction enzymes. Two sites (one for BgIII and one for StuI) were removed from the sequence by replacing codons that were part of the respective recognition sites.

5

10

15

20

25

30

35

Step 5) The sequence generated by steps 1) through 4) was then modified according to allow subsequent generation of a modified NSOPTmut sequence (by homologous recombination). In the sequence obtained from steps 1) through 4) the segment comprising base 3556 to 3755 and the segment comprising base 4456 to 4656 were replaced by the corresponding segments from NSOPTmut. The segment comprising bases 3556 to 4656 of SEQ. ID. NO. 10 can be used to replace the problematic region in NSOPTmut (around position 3900) by homologous recombination thus creating the variant of NSOPTmut having the sequence of SEQ. ID. NO. 11.

Step 6) Analysis of the sequence generated through steps 1) to 5) revealed a potential open reading frame spanning nearly the complete fragment on the complementary strand. Removal of all codons CTA and TTA (Leu) and TCA (Ser) from the sense strand effectively removed all stop codons in one of the reading frames on the complementary strand. Although the likelyhood for transcription of this complementary strand open reading frame and subsequent translation into protein is very small, in order to exclude a potential interference with the transcription and subsequent translation of the sequence encoded on the sense strand, TCA codons for Ser were introduced on the sense approximately every 500 bases. No changes were introduced in the segments introduced during step 5) to allow homologous recombination. The TCA codon for Ser was preferred over the CTA and TTA codons for Leu because of the higher relative frequency for TCA (0.05) as compared to CTA (0.02) and TTA (0.03) in human_high.cod. In addition, the average human codon usage from CUTG favored TCA (0.14 against 0.07 for CTA and TTA).

Step 7) In a final step GCCACC was added at the 5' end of the sequence to generate an optimized internal ribosome entry site (Kozak signal) and a TAAA stop sgnal was added at the 3'. To maintain the initiation of translation

properties of NSsuboptmut the first 8 codons of the coding region were kept identical to the NSOPTmut sequence. The resulting sequence was again checked for the absence of BglII, PmeI and XbaI recognition sites and the presence of only 1 StuI site.

The NSsuboptmut sequence (SEQ. ID. NO. 10) has an overall reduced

GC content (63.5%) as compared to NSOPTmut (70.3%) and maintains a well optimized level of codon usage optimization. Nucleotide sequence identity of NSsuboptmut is 77.2% with respect to NSmut.

Table 5: Definition of codon replacements performed during steps 1) and 2).

•	\sim
Ľ	U

Amino Acid	Most frequent codon	Relative frequency	Reduction in GC content (bases)	Replacement codon	Relative frequency
Amino	Acids where the re	placement codon	reduces the codor	GC-content by 1	base
Ala	GCC	0.51	1	GCT	0.17
Arg	CGC	0.37	1	AGG	0.18
Asn	AAC	0.78	1	AAT	0.22
Asp	GAC	0.75	1	GAT	0.25
Cys	TGC	0.68	1	TGT	0.32
Glu	GAG	0.75	1	GAA	0.25
Gln	CAG	0.88	1	CAA	0.12
Gly	GGC	0.50	1	GGA	0.14
His	CAC	0.79	1	CAT	0.21
Ile	ATC	0.77	1	ATT	0.18
Lys	AAG	0.82	1	AAA	0.18
Phe	TTC	0.80	1	TTT	0.20
Pro	ccc	0.48	1	ССТ	0.19
Ser		0.34	1	TCT	0.13
Thr	AGC	0.51		ACA	0.13
Tyr	ACC		1		0.14
-	TAC	0.74	1	TAT	1 0.26
Mat	Ami	ino Acids with no	alternative codon		T
Met	ATG	1.00	0	ATG	1.00
Trp	TGG	1.00	0	TGG	1.00

Amino Acid	s where the replaceme	nt codon has a very	low relative freq	uency. These amir	no acids were
	excl	uded from the repla	cement procedur	е	
Leu	CTG	0.58	1	TTG	0.06
Val	GTG	0.64	1	GTT	0.07

Example 12: Virus Characterization

Adenovectors were characterized by: (a) measuring the physical particles/ml; (b) running a TaqMan PCR assay; and (c) checking protein expression after infection of HeLa cells.

a) Physical Particles Determination

CsCl purified virus was diluted 1/10 and 1/100 in 0.1% SDS PBS. As a control, buffer A105 was used. These dilutions were incubated 10 minutes at 55° C. After spinning the tubes briefly, O.D. at 260 nm was measured. The amount of viral particles was calculated as follows: 1 OD 260 nm = 1.1×10^{12} physical particles/ml. The results were typically between 5×10^{11} and 1×10^{12} physical particles/ml.

b) TaqMan PCR Assay

10

15

20

25

30

TaqMan PCR assay was used for adenovectors genome quantification (Q-PCR particles/ml). TaqMan PCR assay was performed using the ABI Prism 7700-sequence detector. The reaction was performed in a final 50 μ l volume in the presence of oligonucleotides (at final 200 nM) and probe (at final 200 μ M) specific for the adenoviral backbone. The virus was diluted 1/10 in 0.1% SDS PBS and incubated 10 minutes at 55°C. After spinning the tube briefly, serial 1/10 dilutions (in water) were prepared. 10 μ l the 10°3, 10°5 and 10°7 dilutions were used as templates in the PCR assay.

The amount of particles present in each sample was calculated on the basis of a standard curve run in the same experiment. Typically results were between 1×10^{12} and 3×10^{12} Q-PCR particles/ml.

c) Expression of HCV Non-Structural Proteins

Expression of HCV NS proteins was tested by infection of HeLa cells. Cells were plated the day before the infection at 1.5×10^6 cells/dish (10 cm ø Petri dishes). Different amounts of CsCl purified virus corresponding to m.o.i. of 50, 250

and 1250 pp/cell were diluted in medium (FCS free) up to a final volume of 5 ml. The diluted virus was added on the cells and incubated for 1 hour at 37°C in a CO₂ incubator (gently mixing every 20 minutes). 5 ml of 5% HS-DMEM was added and the cells were incubated at 37°C for 48 hours.

Cell extracts were prepared in 1% Triton/TEN buffer. The extracts were run on 10% SDS-acrylamide gel, blotted on nitrocellulose and assayed with antibodies directed against NS3, NS5a and NS5b in order to check the correct polyprotein cleavage. Mock-infected cells were used as a negative control. Results from representative experiments testing the Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut and MRKAd6-NSOPTmut are shown in Figure 14.

Example 13: Mice Immunization with Adenovectors Encoding Different NS Cassettes

The adenovectors Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut and
MRKAd6-NSOPTmut were injected in C57Black6 mice strains to evaluate their
potential to elicit anti-HCV immune responses. Groups of animals (N=9-10) were
injected intramuscularly with 10⁹ pp of CsCl purified virus. Each animal received two
doses at three weeks interval.

Humoral immune response against the NS3 protein was measured in post dose two sera from C57Black6 immunized mice by ELISA on bacterially expressed NS3 protease domain. Antibodies specific for the tested antigen were detected with geometric mean titers (GMT) ranging from 100 to 46000 (Tables 6, 7, 8 and 9).

Table 6: Ad5-NS

											GMT
Mice n.	1	2	3	4	5	6	7	8	9	10	
Titer	50	253	50	50	50	2257	504	50	50	50	108

30

5

10

20

Table 7: Ad5-NSmut

											GMT
Mice	11	12	13	14	15	16	17	18	19	20	
n. Titer	3162	78850	87241	6796	12134	3340	18473	13093	76167	49593	23645

Table 8: MRKAd6-NSmut

5

20

											GMT
Mice	21	22	23	24	25	26	27	28	29	30	
n. Titer	125626	39751	40187	65834	60619	69933	21555	49348	29290	26859	46461

Table 9: MRKAd6-NSOPTmut

								GMT
Mice n.	31	32	33	34	35	36	37	
Titer	25430	3657	893	175	10442	49540	173	2785

T cell response in C57Black6 mice was analyzed by the quantitative ELISPOT assay measuring the number of IFNγ secreting T cells in response to five pools (named from F to L+M) of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence. Specific CD8+ response induced in C57Black6 mice was analyzed by the same assay using a 20mer peptide encompassing a CD8+ epitope for C57Black6 mice (pep1480). Cells secreting IFNγ in an antigen specific-manner were detected using a standard ELIspot assay.

Spleen cells, splenocytes and peptides were produced and treated as described in Example 3, *supra*. Representative data from groups of C57Black6 mice (N=9-10) immunized with two injections of 10⁹ viral particles of vectors Ad5-NS, MRKAd5-NSmut and MRKAd6-NSmut are shown in Figure 15.

Example 14: Immunization of Rhesus macaques with Adenovectors

Rhesus macaques (N=3-4) were immunized by intramuscular injection of CsCl purified Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut or MRKAd6-

NSOPTmut virus. Each animal received two doses of 10^{11} or 10^{10} vp in the deltoid muscle at 0, and 4 weeks.

CMI was measured at different time points by a) IFN- γ ELISPOT (see Example 3, supra), b) IFN- γ ICS and c) bulk CTL assays. These assays measure HCV antigen-specific CD8+ and CD4+ T lymphocyte responses, and can be used for a variety of mammals, such as humans, rhesus monkeys, mice, and rats.

The use of a specific peptide or a pool of peptides can simplify antigen presentation in CTL cytotoxicity assays, interferon-gamma ELISPOT assays and interferon-gamma intracellular staining assays. Peptides based on the amino acid sequence of various HCV proteins (core, E2, NS3, NS4A, NS4B, NS5a, NS5b) were prepared for use in these assays to measure immune responses in HCV DNA and adenovirus vector vaccinated rhesus monkeys, as well as in HCV-infected humans. The individual peptides are overlapping 20-mers, offset by 10 amino acids. Large pools of peptides can be used to detect an overall response to HCV proteins while smaller pools and individual peptides may be used to define the epitope specificity of a response.

IFN-γICS

5

10

15

20

25

30

35

For IFN- γ ICS, 2 x 106 PBMC in 1 ml R10 (RPMI medium, supplemented with 10% FCS) were stimulated with peptide pool antigens. Final concentration of each peptide was 2 μ g/ml. Cells were incubated for 1 hour in a CO₂ incubator at 37°C and then Brefeldin A was added to a final concentration of 10 μ g /ml to inhibit the secretion of soluble cytokines. Cells were incubated for additional 14-16 hours at 37°C.

Stimulation was done in the presence of co-stimulatory antibodies: CD28 and CD49d (anti-humanCD28 BD340975 and anti-humanCD49d BD340976). After incubation, cells were stained with fluorochrome-conjugated antibodies for surface antigens: anti-CD3, anti-CD4, anti-CD8 (CD3-APC Biosource APS0301, CD4-PE BD345769, CD8-PerCP BD345774).

To detect intracellular cytokines, cells were treated with FACS permeabilization buffer 2 (BD340973), 2x final concentration. Once fixed and permeabilized, cells were incubated with an antibody against human IFN-γ, IFN-γFTC (Biosource AHC4338).

Cells were resuspended in 1% formaldehyde in PBS and analyzed at FACS within 24 hours. Four color FACS analysis was performed on a FACSCalibur

instrument (Becton Dickinson) equipped with two lasers. Acquisition was done gating on the lymphocyte population in the Forward versus Side Scatter plot coupled with the CD3, CD8 positive populations. At least 30,000 events of the gate were taken. The positive cells are expressed as number of IFN- γ expressing cells over 10^6 lymphocytes.

IFN- γ ELISPOT and IFN- γ ICS data from immunized monkeys after one or two injections of 10^{10} or 10^{11} vp of the different adenovectors are reported in Figures 16A-16D, 17A, and 17B.

10 Bulk CTL Assays

5

15

20

25

30

A distinguishing effector function of T lymphocytes is the ability of subsets of this cell population to directly lyse cells exhibiting appropriate MHC-associated antigenic peptides. This cytotoxic activity is most often associated with CD8+ T lymphocytes.

PBMC samples were infected with recombinant vaccine viruses expressing HCV antigens *in vitro* for approximately 14 days to provide antigen restimulation and expansion of memory T cells. Cytotoxicity against autologous B cell lines treated with peptide antigen pools was tested.

The lytic function of the culture is measured as a percentage of specific lysis resulted from chromium released from target cells during 4 hours incubation with CTL effector cells. Specific cytotoxicity is measured and compared to irrelevant antigen or excipient-treated B cell lines. This assay is semi-quantitative and is the preferred means for determining whether CTL responses were elicited by the vaccine. Data after two injections from monkeys immunized with 10¹¹ vp/dose with adenovectors Ad5-NS, MRKAd5-NSmut and MRKAd6-NSmut are reported in Figures 18A-18F.

Other embodiments are within the following claims. While several embodiments have been shown and described, various modifications may be made without departing from the spirit and scope of the present invention.

WHAT IS CLAIMED IS:

5

10

20

30

1. A nucleic acid comprising a nucleotide sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ ID NO: 1, provided that said polypeptide has sufficient protease activity to process itself to produce an NS5B protein and said NS5B protein is enzymatically inactive.

- 2. The nucleic acid of claim 1, wherein said nucleotide sequence is substantially similar to the coding sequence of SEQ ID NO: 2.
- 3. The nucleic acid of claim 1, wherein said nucleotide sequence encodes for the polypeptide of SEQ ID NO: 1.
- 4. The nucleic acid of claim 3, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
 - 5. The nucleic acid of claim 3, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2 or SEQ ID NO: 3.
 - 6. The nucleic acid of any one of claims 1-5, wherein said nucleic acid is an expression vector capable of expressing said polypeptide from said nucleotide sequence in a human cell.
- 7. A nucleic acid comprising a gene expression cassette able to express a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ ID NO: 1 in a human cell, provided that said polypeptide can process itself to produce an NS5B protein and said NS5B protein is enzymatically inactive, said expression cassette comprising:
 - a) a promoter transcriptionally coupled to a nucleotide sequence encoding said polypeptide;
 - b) a 5' ribosome binding site functionally coupled to said nucleotide sequence,

c) a terminator joined to the 3' end of said nucleotide sequence, and
d) a 3' polyadenylation signal functionally coupled to said nucleotide sequence.

- 5 8. The nucleic acid of claim 7, wherein said nucleotide sequence is substantially similar to either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
- 9. The nucleic acid of claim 8, wherein said nucleic acid is a shuttle vector further comprising a selectable marker, an origin of replication, a first adenovirus homology region and a second adenovirus homology region flanking said expression cassette, wherein said first homology region has at least about 100 base pairs substantially homologous to at least right end of a wild-type adenovirus region from about base pairs 1-425, and said second homology region has at least about 100 base pairs substantially homologous to at least the left end of a wild-type adenovirus region from about base pairs 3511-5792 of Ad5 or corresponding region of another adenovirus.
- 10. The nucleic acid of claim 9, wherein said nucleotide sequence 20 encodes for a polypeptide of SEQ ID NO: 1.
 - 11. The nucleic acid of claim 9, wherein said nucleotide sequence is SEQ ID NO: 2.
- 25 12. The nucleic acid of claim 9, wherein said nucleotide sequence is either SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
 - 13. The nucleic acid of claim 8, wherein said nucleic acid is a plasmid suitable for administration into a human and further comprises a prokaryotic origin of replication and a gene coding for a selectable marker.

30

14. The nucleic acid of claim 13, wherein said nucleotide sequence encodes for a polypeptide of SEQ ID NO: 1.

15. The nucleic acid of claim 14, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

- 5 16. The nucleic acid of claim 14, wherein said nucleotide sequence is the coding sequence of SEQ ID NO: 2 or SEQ ID NO: 3.
 - 17. The nucleic acid of claim 14, wherein said promoter is the human intermediate early cytomegalovirus promoter (intron A), said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the bovine growth hormone (BGH) polyadenylation signal.

10

20

25

- 18. The nucleic acid of claim 8, wherein said nucleic acid is a adenovirus genome plasmid comprising a selectable marker, an origin of replication, and a recombinant adenovector genome containing an E1 deletion, an E3 deletion, and said expression cassette.
 - 19. The nucleic acid of claim 8, wherein said nucleic acid is a adenovirus genome plasmid comprising a selectable marker, an origin of replication, and
 - a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) said gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to said first region;
 - c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
 - e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and

f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

- 5 20. The nucleic acid of claim 19, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.
- 10 21. The nucleic acid of claim 20, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.
- 15 22. The nucleic acid of claim 21, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
- 23. The nucleic acid of claim 19, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.
- 24. The nucleic acid of claim 23, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.
- 25. The nucleic acid of claim 24, wherein said expression cassette
 30 is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO:
 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
- 26. The nucleic acid of claim 24, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO:
 2 or SEQ ID NO: 3.

27. The nucleic acid of claim 8, wherein said nucleic acid is a adenovirus genome plasmid comprising an origin of replication, a selectable marker, and:

a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

5

10

15

20

25

30

- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.
- 28. The nucleic acid of claim 27, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.
- 29. The nucleic acid of claim 28, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.
- 30. The nucleic acid of claim 27, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 of Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

31. The nucleic acid of claim 30, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

5

10

20

25

30

- 32. The nucleic acid of claim 8, wherein said nucleic acid is a adenovector consisting of a nucleotide sequence substantially similar to of SEQ ID NO. 4 or a derivative thereof, wherein said derivative thereof has the HCV polyprotein encoding sequence present in SEQ ID NO: 4 replaced with the HCV polyprotein encoding sequence of either SEQ ID NO: 3, SEQ ID NO: 10 or SEQ ID NO: 11.
- 33. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector having an adenovector genome containing an E1 deletion, an E3 deletion, and said expression cassette
 - 34. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector consisting of:
 - a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) said gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to said first region;
 - c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
 - e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

35. The nucleic acid of claim 34, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

- 36. The nucleic acid of claim 35, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.
- 37. The nucleic acid of claim 36, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

15

10

5

38. The nucleic acid of claim 34, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

- 39. The nucleic acid of claim 37, where said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.
- 25 40. The nucleic acid of claim 39, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
- 41. The nucleic acid of claim 39, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is SEQ ID NO: 2 or SEQ ID NO: 3.
 - 42. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector consisting of:

a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;

5

- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel
 orientation joined to said third region;
 - e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.
- 43. The nucleic acid of claim 42, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.
 - 44. The nucleic acid of claim 42, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.
- ID NO. 4 or a derivative thereof, wherein said derivative thereof has the HCV polyprotein encoding sequence present in SEQ ID NO: 4 replaced with the HCV polyprotein encoding sequence of either SEQ ID NO: 3, SEQ ID NO: 10 or SEQ ID NO: 11.
 - 46. An adenovector produced by a process comprising the steps of:

a) producing an adenovirus genome plasmid by homologous recombination between the shuttle vector of claim 9 and a nucleic acid comprising; a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;

5

10

15

a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;

a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and

a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region; and

- b) rescuing said adenovector from said adenovirus plasmid.
- 47. A cultured recombinant cell comprising the nucleic acid of 20 claim 6.
 - 48. A cultured recombinant cell comprising the nucleic acid of any one of claims 9-46.
- 25 49. A method of making an adenovector comprising the steps of:
 - a) producing an adenovirus genome plasmid comprising a gene
 expression cassette by homologous recombination between the nucleic acid of claim 9
 and a nucleic acid comprising;

a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;

a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;

- a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and
 - a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region; and
 - b) rescuing said recombinant adenovirus from said recombinant adenovirus plasmid.
 - 50. A pharmaceutical composition comprising the nucleic acid of any one of claims 13-17 and 32-46 and pharmaceutically acceptable carrier.
 - 51. A method of treating a patient comprising the step of administering to said patient an effective amount of the nucleic acid of any one of claims 13-17 and 32-46.
- 20 52. The method of claim 51, wherein said patient is a human.
 - 53. The method of claim 52, wherein said patient is not infected with HCV.
- The method of claim 52, wherein said patient is infected with HCV.
 - 55. A recombinant nucleic acid comprising one or more Ad6 regions and a region not present in Ad6, wherein at least one Ad6 region is selected from the group consisting of: E1A, E1B, E2B, E2A, E4, L1, L2, L4, and L5.
 - 56. The recombinant nucleic acid of claim 55, wherein said region not present in Ad6, is an expression cassette coding for a polypeptide not found in Ad6.

35

30

10

57. The recombinant nucleic acid of claim 56, wherein said recombinant nucleic acid is an adenovirus vector defective in at least E1 that is able to replicate when E1 is supplied *in trans*.

- 58. The recombinant nucleic acid of claim 57, wherein said vector consists of:
 - a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) said gene expression cassette in an E1 parallel or E1 antiparallel orientation joined to said first region;
 - c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said gene expression cassette;
- d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
 - e) an optionally present fourth region from about base pair 28134 to about base pair 30817 corresponding to Ad5, or from about base pair 28157 to about 30789 corresponding to Ad6, joined to said third region;
 - f) a fifth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, wherein said fifth region is joined to said fourth region if said fourth region is present, or said fifth is joined to said third region if said fourth region is not present; and
 - g) a sixth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region;

provided that at least one of said second, third, and fifth regions is from Ad6.

30

5

10

15

20

25

- 59. The recombinant nucleic acid of claim 57, wherein said vector consists of:
- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;

- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
 - d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region;
- provided that at least one of said second, third, and fourth regions is from Ad6.

1	MAPITAYSQQ	TRGLLGCIIT	SLTGRDKNQV	EGEVQVVSTA	TQSFLATCVN
51	GVCWTVYHGA	GSKTLAGPKG	PITQMYTNVD	QDLVGWQAPP	GARSLTPCTC
101	GSSDLYLVTR	HADVIPVRRR	GDSRGSLLSP	RPVSYLKGSS	GGPLLCPSGH
151	AVGIFRAAVC	TRGVAKAVDF	VPVESMETTM	RSPVFTDNSS	PPAVPQSFQV
201	AHLHAPTGSG	KSTKVPAAYA	AQGYKVLVLN	PSVAATLGFG	AYMSKAHGID
251	PNIRTGVRTI	TTGAPVTYST	YGKFLADGGC	SGGAYDIIIC	DECHSTDSTT
301	ILGIGTVLDQ	AETAGARLVV	LATATPPGSV	TVPHPNIEEV	ALSNTGEIPF
351	YGKAIPIEAI	RGGRHLIFCH	SKKKCDELAA	KLSGLGINAV	AYYRGLDVSV
401	IPTIGDVVVV	ATDALMTGYT	GDFDSVIDCN	TCVTQTVDFS	LDPTFTIETT
451	TVPQDAVSRS	QRRGRTGRGR	RGIYRFVTPG	ERPSGMFDSS	VLCECYDAGC
501	AWYELTPAET	SVRLRAYLNT	PGLPVCQDHL	EFWESVFTGL	THIDAHFLSQ
551	${\tt TKQAGDNFPY}$	LVAYQATVCA	RAQAPPPSWD	QMWKCLIRLK	PTLHGPTPLL
601	YRLGAVQNEV	TLTHPITKYI	MACMSADLEV	VTSTWVLVGG	VLAALAAYCL
651	TTGSVVIVGR	IILSGRPAIV	PDREFLYQEF	DEMEECASHL	PYIEQGMQLA
701	EQFKQKALGL	LQTATKQAEA	AAPVVESKWR	ALETFWAKHM	WNFISGIQYL
75 1	AGLSTLPGNP	AIASLMAFTA	SITSPLTTQS	TLLFNILGGW	VAAQLAPPSA
801	ASAFVGAGIA	GAAVGSIGLG	KVLVDILAGY	GAGVAGALVA	FKVMSGEMPS
851	TEDLVNLLPA	ILSPGALVVG	VVCAAILRRH	VGPGEGAVQW	MNRLIAFASR
901	GNHVSPTHYV	PESDAAARVT	QILSSLTITQ	LLKRLHQWIN	EDCSTPCSGS
951	WLRDVWDWIC	TVLTDFKTWL	QSKLLPQLPG	VPFFSCQRGY	KGVWRGDGIM
1001	QTTCPCGAQI	TGHVKNGSMR	IVGPKTCSNT	WHGTFPINAY	TTGPCTPSPA
1051	PNYSRALWRV	AAEEYVEVTR	VGDFHYVTGM	TTDNVKCPCQ	VPAPEFFTEV
1101	DGVRLHRYAP	ACRPLLREEV	TFQVGLNQYL	VGSQLPCEPE	PDVAVLTSML
1151	TDPSHITAET	AKRRLARGSP	PSLASSSASQ	LSAPSLKATC	TTHHVSPDAD
1201	LIEANLLWRQ	EMGGNITRVE	SENKVVVLDS	FDPLRAEEDE	REVSVPAEIL
1251	RKSKKFPAAM	PIWARPDYNP	PLLESWKDPD	YVPPVVHGCP	LPPIKAPPIP
1301	PPRRKRTVVL	TESSVSSALA	ELATKTFGSS	ESSAVDSGTA	TALPDQASDD
1351	GDKGSDVESY	SSMPPLEGEP	GDPDLSDGSW	STVSEEASED	VVCCSMSYTW
1401	TGALITPCAA	EESKLPINAL	SNSLLRHHNM	VYATTSRSAG	LRQKKVTFDR
1451	LQVLDDHYRD	VLKEMKAKAS	TVKAKLLSVE	EACKLTPPHS	AKSKFGYGAK
1501	DVRNLSSKAV	NHIHSVWKDL	LEDTVTPIDT	TIMAKNEVFC	VQPEKGGRKP
1551	ARLIVFPDLG	VRVCEKMALY	DVVSTLPQVV	MGSSYGFQYS	PGQRVEFLVN
1601	TWKSKKNPMG	FSYDTRCFDS	TVTENDIRVE	ESIYQCCDLA	PEARQAIKSL
1651	TERLYIGGPL	TNSKGQNCGY	RRCRASGVLT	TSCGNTLTCY	LKASAACRAA

FIG. 1A

1701	KLQDCTMLVN	AAGLVVICES	AGTQEDAASL	RVFTEAMTRY	SAPPGDPPQP
1751	EYDLELITSC	SSNVSVAHDA	SGKRVYYLTR	DPTTPLARAA	WETARHTPVN
1801	SWLGNIIMYA	PTLWARMILM	THFFSILLAQ	EQLEKALDCQ	IYGACYSIEP
1851	LDLPQIIERL	HGLSAFSLHS	YSPGEINRVA	SCLRKLGVPP	LRVWRHRARS
1901	VRARLLSQGG	RAATCGKYLF	NWAVKTKLKL	TPIPAASQLD	LSGWFVAGYS
1951	GGDIYHSLSR	ARPRWFMLCL	LLLSVGVGIY	LLPNR	

1	GCCACCATGG	CGCCCATCAC	GGCCTACTCC	CAACAGACGC	GGGGCCTACT
51	TGGTTGCATC	ATCACTAGCC	TTACAGGCCG	GGACAAGAAC	CAGGTCGAGG
101	GAGAGGTTCA	GGTGGTTTCC	ACCGCAACAC	AATCCTTCCT	GGCGACCTGC
151	GTCAACGGCG	TGTGTTGGAC	CGTTTACCAT	GGTGCTGGCT	CAAAGACCTT
201	AGCCGGCCCA	AAGGGCCAA	TCACCCAGAT	GTACACTAAT	GTGGACCAGG
251	ACCTCGTCGG	CTGGCAGGCG	cccccggg	CGCGTTCCTT	GACACCATGC
301	ACCTGTGGCA	GCTCAGACCT	TTACTTGGTC	ACGAGACATG	CTGACGTCAT
351	TCCGGTGCGC	CGGCGGGCG	ACAGTAGGGG	GAGCCTGCTC	TCCCCCAGGC
401	CTGTCTCCTA	CTTGAAGGGC	TCTTCGGGTG	GTCCACTGCT	CTGCCCTTCG
451	GGGCACGCTG	TGGGCATCTT	CCGGGCTGCC	GTATGCACCC	GGGGGGTTGC
501	GAAGGCGGTG	GACTTTGTGC	CCGTAGAGTC	CATGGAAACT	ACTATGCGGT
551	CTCCGGTCTT	CACGGACAAC	TCATCCCCCC	CGGCCGTACC	GCAGTCATTT
601	CAAGTGGCCC	ACCTACACGC	TCCCACTGGC	AGCGGCAAGA	GTACTAAAGT
651	GCCGGCTGCA	TATGCAGCCC	AAGGGTACAA	GGTGCTCGTC	CTCAATCCGT
701	CCGTTGCCGC	TACCTTAGGG	TTTGGGGCGT	ATATGTCTAA	GGCACACGGT
751	ATTGACCCCA	ACATCAGAAC	TGGGGTAAGG	ACCATTACCA	CAGGCGCCCC
801	CGTCACATAC	TCTACCTATG	GCAAGTTTCT	TGCCGATGGT	GGTTGCTCTG
851	GGGGCGCTTA	TGACATCATA	ATATGTGATG	AGTGCCATTC	AACTGACTCG
901	ACTACAATCT	TGGGCATCGG	CACAGTCCTG	GACCAAGCGG	AGACGGCTGG
951	AGCGCGGCTT	GTCGTGCTCG	CCACCGCTAC	GCCTCCGGGA	TCGGTCACCG
1001	TGCCACACCC	AAACATCGAG	GAGGTGGCCC	TGTCTAATAC	TGGAGAGATC
1051	CCCTTCTATG	GCAAAGCCAT	CCCCATTGAA	GCCATCAGGG	GGGGAAGGCA
1101	TCTCATTTTC	TGTCATTCCA	AGAAGAAGTG	CGACGAGCTC	GCCGCAAAGC
1151	TGTCAGGCCT	CGGAATCAAC	GCTGTGGCGT	ATTACCGGGG	GCTCGATGTG
1201	TCCGTCATAC	CAACTATCGG	AGACGTCGTT	GTCGTGGCAA	CAGACGCTCT
1251	GATGACGGGC	TATACGGGCG	ACTTTGACTC	AGTGATCGAC	TGTAACACAT
1301	GTGTCACCCA	GACAGTCGAC	TTCAGCTTGG	ATCCCACCTT	CACCATTGAG
1351	ACGACGACCG	TGCCTCAAGA	CGCAGTGTCG	CGCTCGCAGC	GGCGGGGTAG
1401	GACTGGCAGG	GGTAGGAGAG	GCATCTACAG	GTTTGTGACT	CCGGGAGAAC
1451	GGCCCTCGGG	CATGTTCGAT	TCCTCGGTCC	TGTGTGAGTG	CTATGACGCG
1501	GGCTGTGCTT	GGTACGAGCT	CACCCCCGCC	GAGACCTCGG	TTAGGTTGCG
1551	GGCCTACCTG	AACACACCAG	GGTTGCCCGT	TTGCCAGGAC	CACCTGGAGT
1601	TCTGGGAGAG	TGTCTTCACA	GGCCTCACCC	ACATAGATGC	ACACTTCTTG
1651	TCCCAGACCA	AGCAGGCAGG	AGACAACTTC	CCCTACCTGG	TAGCATACCA

					mada a maa a a
1701		TGCGCCAGGG			
1751		TCTCATACGG			
1801		GGCTGGGAGC			
1851		TACATCATGG			
1901	CTAGCACCTG	GGTGCTGGTG	GGCGGAGTCC	TTGCAGCTCT	GGCCGCGTAT
1951	TGCCTGACAA	CAGGCAGTGT	GGTCATTGTG	GGTAGGATTA	TCTTGTCCGG
2001	GAGGCCGGCT	ATTGTTCCCG	ACAGGGAGTT	TCTCTACCAG	GAGTTCGATG
2051	AAATGGAAGA	GTGCGCCTCG	CACCTCCCTT	ACATCGAGCA	GGGAATGCAG
2101	CTCGCCGAGC	AATTCAAGCA	GAAAGCGCTC	GGGTTACTGC	AAACAGCCAC
2151	CAAACAAGCG	GAGGCTGCTG	CTCCCGTGGT	GGAGTCCAAG	TGGCGAGCCC
2201	TTGAGACATT	CTGGGCGAAG	CACATGTGGA	ATTTCATCAG	CGGGATACAG
2251	TACTTAGCAG	GCTTATCCAC	TCTGCCTGGG	AACCCCGCAA	TAGCATCATT
2301	GATGGCATTC	ACAGCCTCTA	TCACCAGCCC	GCTCACCACC	CAAAGTACCC
2351	TCCTGTTTAA	CATCTTGGGG	GGGTGGGTGG	CTGCCCAACT	CGCCCCCCC
2401	AGCGCCGCTT	CGGCTTTCGT	GGGCGCCGGC	ATCGCCGGTG	CGGCTGTTGG
2451	CAGCATAGGC	CTTGGGAAGG	TGCTTGTGGA	CATTCTGGCG	GGTTATGGAG
2501	CAGGAGTGGC	CGGCGCGCTC	GTGGCCTTCA	AGGTCATGAG	CGGCGAGATG
2551	CCCTCCACCG	AGGACCTGGT	CAATCTACTT	CCTGCCATCC	TCTCTCCTGG
2601	CGCCCTGGTC	GTCGGGGTCG	TGTGTGCAGC	AATACTGCGT	CGACACGTGG
2651	GTCCGGGAGA	GGGGGCTGTG	CAGTGGATGA	ACCGGCTGAT	AGCGTTCGCC
2701	TCGCGGGGTA	ATCATGTTTC	CCCCACGCAC	TATGTGCCTG	AGAGCGACGC
2751	CGCAGCGCGT	GTTACTCAGA	TCCTCTCCAG	CCTTACCATC	ACTCAGCTGC
2801	TGAAAAGGCT	CCACCAGTGG	ATTAATGAAG	ACTGCTCCAC	ACCGTGTTCC
2851	GGCTCGTGGC	TAAGGGATGT	TTGGGACTGG	ATATGCACGG	TGTTGACTGA
2901 ·	CTTCAAGACC	TGGCTCCAGT	CCAAGCTCCT	GCCGCAGCTA	CCGGGAGTCC
2951	CTTTTTTCTC	GTGCCAACGC	GGGTACAAGG	GAGTCTGGCG	GGGAGACGGC
3001	ATCATGCAAA	CCACCTGCCC	ATGTGGAGCA	CAGATCACCG	GACATGTCAA
3051	AAACGGTTCC	ATGAGGATCG	TCGGGCCTAA	GACCTGCAGC	AACACGTGGC
3101	ATGGAACATT	CCCCATCAAC	GCATACACCA	CGGGCCCCTG	CACACCCTCT
3151	CCAGCGCCAA	ACTATTCTAG	GGCGCTGTGG	CGGGTGGCCG	CTGAGGAGTA
3201	CGTGGAGGTC	ACGCGGGTGG	GGGATTTCCA	CTACGTGACG	GGCATGACCA
3251	CTGACAACGT	AAAGTGCCCA	TGCCAGGTTC	CGGCTCCTGA	ATTCTTCACG
3301					GCAGGCCTCT
3351					TACCTGGTTG

FIG. 2B

3401	GGTCACAGCT	ACCATGCGAG	CCCGAACCGG	ATGTAGCAGT	GCTCACTTCC
3451	ATGCTCACCG	ACCCCTCCCA	CATCACAGCA	GAAACGGCTA	AGCGTAGGTT
3501	GGCCAGGGGG	TCTCCCCCT	CCTTGGCCAG	CTCTTCAGCT	AGCCAGTTGT
3551	CTGCGCCTTC	CTTGAAGGCG	ACATGCACTA	CCCACCATGT	CTCTCCGGAC
3601	GCTGACCTCA	TCGAGGCCAA	CCTCCTGTGG	CGGCAGGAGA	TGGGCGGGAA
3651	CATCACCCGC	GTGGAGTCGG	AGAACAAGGT	GGTAGTCCTG	GACTCTTTCG
3701	ACCCGCTTCG	AGCGGAGGAG	GATGAGAGGG	AAGTATCCGT	TCCGGCGGAG
3751	ATCCTGCGGA	AATCCAAGAA	GTTCCCCGCA	GCGATGCCCA	TCTGGGCGCG
3801	CCCGGATTAC	AACCCTCCAC	TGTTAGAGTC	CTGGAAGGAC	CCGGACTACG
3851	TCCCTCCGGT	GGTGCACGGG	TGCCCGTTGC	CACCTATCAA	GGCCCCTCCA
3901	ATACCACCTC	CACGGAGAAA	GAGGACGGTT	GTCCTAACAG	AGTCCTCCGT
3951	GTCTTCTGCC	TTAGCGGAGC	TCGCTACTAA	GACCTTCGGC	AGCTCCGAAT
4001	CATCGGCCGT	CGACAGCGGC	ACGGCGACCG	CCCTTCCTGA	CCAGGCCTCC
4051	GACGACGGTG	ACAAAGGATC	CGACGTTGAG	TCGTACTCCT	CCATGCCCCC
4101	CCTTGAGGGG	GAACCGGGGG	ACCCCGATCT	CAGTGACGGG	TCTTGGTCTA
4151	CCGTGAGCGA	GGAAGCTAGT	GAGGATGTCG	TCTGCTGCTC	AATGTCCTAC
4201	ACATGGACAG	${\tt GCGCCTTGAT}$	CACGCCATGC	GCTGCGGAGG	AAAGCAAGCT
4251	GCCCATCAAC	GCGTTGAGCA	ACTCTTTGCT	GCGCCACCAT	AACATGGTTT
4301	ATGCCACAAC	ATCTCGCAGC	GCAGGCCTGC	GGCAGAAGAA	GGTCACCTTT
4351	GACAGACTGC	AAGTCCTGGA	CGACCACTAC	CGGGACGTGC	TCAAGGAGAT
4401	GAAGGCGAAG	GCGTCCACAG	TTAAGGCTAA	ACTCCTATCC	GTAGAGGAAG
4451	CCTGCAAGCT	GACGCCCCCA	CATTCGGCCA	AATCCAAGTT	TGGCTATGGG
4501	GCAAAGGACG	TCCGGAACCT	ATCCAGCAAG	GCCGTTAACC	ACATCCACTC
4551	CGTGTGGAAG	GACTTGCTGG	AAGACACTGT	GACACCAATT	GACACCACCA
4601	TCATGGCAAA	AAATGAGGTT	TTCTGTGTCC	AACCAGAGAA	AGGAGGCCGT
4651	AAGCCAGCCC	GCCTTATCGT	ATTCCCAGAT	CTGGGAGTCC	GTGTATGCGA
4701	GAAGATGGCC	CTCTATGATG	TGGTCTCCAC	CCTTCCTCAG	GTCGTGATGG
4751	GCTCCTCATA	CGGATTCCAG	TACTCTCCTG	GGCAGCGAGT	CGAGTTCCTG
4801	GTGAATACCT	GGAAATCAAA	GAAAAACCCC	ATGGGCTTTT	CATATGACAC
4851	TCGCTGTTTC	GACTCAACGG	TCACCGAGAA	CGACATCCGT	GTTGAGGAGT
4901	CAATTTACCA	ATGTTGTGAC	TTGGCCCCCG	AAGCCAGACA	GGCCATAAAA
4951	TCGCTCACAG	AGCGGCTTTA	TATCGGGGGT	CCTCTGACTA	ATTCAAAAGG
5001	GCAGAACTGC	GGTTATCGCC	GGTGCCGCGC	GAGCGGCGTG	CTGACGACTA
5051	GCTGCGGTAA	CACCCTCACA	TGTTACTTGA	AGGCCTCTGC	AGCCTGTCGA

5101	GCTGCGAAGC	TCCAGGACTG	CACGATGCTC	GTGAACGCCG	CCGGCCTTGT
5151	CGTTATCTGT			GGACGCGGCG	
5201	TCTTCACGGA	GGCTATGACT	AGGTACTCTG	CCCCCCCGG	GGACCCGCCC
5251	CAACCAGAAT	ACGACTTGGA	GCTGATAACA	TCATGTTCCT	CCAATGTGTC
5301	GGTCGCCCAC	GATGCATCAG	GCAAAAGGGT	GTACTACCTC	ACCCGTGATC
5351	CCACCACCCC	CCTCGCACGG	GCTGCGTGGG	AAACAGCTAG	ACACACTCCA
5401	GTTAACTCCT	GGCTAGGCAA	CATTATCATG	TATGCGCCCA	CTTTGTGGGC
5451	AAGGATGATT	CTGATGACTC	ACTTCTTCTC	CATCCTTCTA	GCACAGGAGC
5501	AACTTGAAAA	AGCCCTGGAC	TGCCAGATCT	ACGGGGCCTG	TTACTCCATT
5551	GAGCCACTTG	ACCTACCTCA	GATCATTGAA	CGACTCCATG	GCCTTAGCGC
5601	ATTTTCACTC	CATAGTTACT	CTCCAGGTGA	GATCAATAGG	GTGGCTTCAT
5651	GCCTCAGGAA	ACTTGGGGTA	CCACCCTTGC	GAGTCTGGAG	ACATCGGGCC
5701	AGGAGCGTCC	GCGCTAGGCT	ACTGTCCCAG	GGGGGAGGG	CCGCCACTTG
5751	TGGCAAGTAC	CTCTTCAACT	GGGCAGTGAA	GACCAAACTC	AAACTCACTC
5801	CAATCCCGGC	TGCGTCCCAG	CTGGACTTGT	CCGGCTGGTT	CGTTGCTGGT
5851	TACAGCGGGG	GAGACATATA	TCACAGCCTG	TCTCGTGCCC	GACCCCGCTG
5901	GTTCATGCTG	TGCCTACTCC	TACTTTCTGT	AGGGGTAGGC	ATCTACCTGC
5951	TCCCCAACCG	ATAAA			

1	GCCACCATGG	CCCCCATCAC	CGCCTACAGC	CAGCAGACCC	GCGGCCTGCT
51	GGGCTGCATC	ATCACCAGCC	TGACCGGCCG	CGACAAGAAC	CAGGTGGAGG
101	GCGAGGTGCA	GGTGGTGAGC	ACCGCCACCC	AGAGCTTCCT	GGCCACCTGC
151	GTGAACGGCG	TGTGCTGGAC	CGTGTACCAC	GGCGCCGGCA	GCAAGACCCT
201	GGCCGGCCCC	AAGGGCCCCA	TCACCCAGAT	GTACACCAAC	GTGGACCAGG
251	ACCTGGTGGG	CTGGCAGGCC	CCCCCGGCG	CCCGCAGCCT	GACCCCCTGC
301	ACCTGCGGCA	GCAGCGACCT	GTACCTGGTG	ACCCGCCACG	CCGACGTGAT
351	CCCCGTGCGC	CGCCGCGGCG	ACAGCCGCGG	CAGCCTGCTG	AGCCCCCGCC
401	CCGTGAGCTA	CCTGAAGGGC	AGCAGCGGCG	GCCCCTGCT	GTGCCCCAGC
451	GGCCACGCCG	TGGGCATCTT	CCGCGCCGCC	GTGTGCACCC	GCGGCGTGGC
501	CAAGGCCGTG	GACTTCGTGC	CCGTGGAGAG	CATGGAGACC	ACCATGCGCA
551	GCCCCGTGTT	CACCGACAAC	AGCAGCCCCC	CCGCCGTGCC	CCAGAGCTTC
601	CAGGTGGCCC	ACCTGCACGC	CCCCACCGGC	AGCGGCAAGA	GCACCAAGGT
651	GCCCGCCGCC	TACGCCGCCC	AGGGCTACAA	GGTGCTGGTG	CTGAACCCCA
701	GCGTGGCCGC	CACCCTGGGC	TTCGGCGCCT	ACATGAGCAA	GGCCCACGGC
751	ATCGACCCCA	ACATCCGCAC	CGGCGTGCGC	ACCATCACCA	CCGGCGCCCC
801	CGTGACCTAC	AGCACCTACG	GCAAGTTCCT	GGCCGACGGC	GGCTGCAGCG
851	GCGGCGCCTA	CGACATCATC	ATCTGCGACG	AGTGCCACAG	CACCGACAGC
901	ACCACCATCC	TGGGCATCGG	CACCGTGCTG	GACCAGGCCG	AGACCGCCGG
951	CGCCCGCCTG	GTGGTGCTGG	CCACCGCCAC	CCCCCCGGC	AGCGTGACCG
1001	TGCCCCACCC	CAACATCGAG	GAGGTGGCCC	TGAGCAACAC	CGGCGAGATC
1051	CCCTTCTACG	GCAAGGCCAT	CCCCATCGAG	GCCATCCGCG	GCGGCCGCCA
1101	CCTGATCTTC	TGCCACAGCA	AGAAGAAGTG	CGACGAGCTG	GCCGCCAAGC
1151	TGAGCGGCCT	GGGCATCAAC	GCCGTGGCCT	ACTACCGCGG	CCTGGACGTG
1201	AGCGTGATCC	CCACCATCGG	CGACGTGGTG	GTGGTGGCCA	CCGACGCCCT
1251	GATGACCGGC	TACACCGGCG	ACTTCGACAG	CGTGATCGAC	TGCAACACCT
1301	GCGTGACCCA	GACCGTGGAC	TTCAGCCTGG	ACCCCACCTT	CACCATCGAG
1351	ACCACCACCG	TGCCCCAGGA	CGCCGTGAGC	CGCAGCCAGC	GCCGCGGCCG
1401	CACCGGCCGC	GGCCGCCGCG	GCATCTACCG	CTTCGTGACC	CCCGGCGAGC
1451	GCCCAGCGG	CATGTTCGAC	AGCAGCGTGC	TGTGCGAGTG	CTACGACGCC
1501	GGCTGCGCCT	GGTACGAGCT	GACCCCCGCC	GAGACCAGCG	TGCGCCTGCG
1551	CGCCTACCTG	AACACCCCCG	GCCTGCCCGT	GTGCCAGGAC	CACCTGGAGT
1601	TCTGGGAGAG	CGTGTTCACC	GGCCTGACCC	ACATCGACGC	CCACTTCCTG
1651	AGCCAGACCA	AGCAGGCCGG	CGACAACTTC	CCCTACCTGG	TGGCCTACCA

FIG. 3A

1701			CCCAGGCCCC		
1751			CTGAAGCCCA		
1801			CGTGCAGAAC		
1851	CATCACCAAG	TACATCATGG	CCTGCATGAG	CGCCGACCTG	GAGGTGGTGA
1901			GGCGGCGTGC		
1951	TGCCTGACCA	CCGGCAGCGT	GGTGATCGTG	GGCCGCATCA	TCCTGAGCGG
2001	CCGCCCCGCC	ATCGTGCCCG	ACCGCGAGTT	CCTGTACCAG	GAGTTCGACG
2051	AGATGGAGGA	GTGCGCCAGC	CACCTGCCCT	ACATCGAGCA	GGGCATGCAG
2101	CTGGCCGAGC	AGTTCAAGCA	GAAGGCCCTG	GGCCTGCTGC	AGACCGCCAC
2151	CAAGCAGGCC	GAGGCCGCCG	CCCCGTGGT	GGAGAGCAAG	TGGCGCGCCC
2201	TGGAGACCTT	CTGGGCCAAG	CACATGTGGA	ACTTCATCAG	CGGCATCCAG
2251	TACCTGGCCG	GCCTGAGCAC	CCTGCCCGGC	AACCCCGCCA	TCGCCAGCCT
2301	GATGGCCTTC	ACCGCCAGCA	TCACCAGCCC	CCTGACCACC	CAGAGCACCC
2351	TGCTGTTCAA	CATCCTGGGC	GGCTGGGTGG	CCGCCCAGCT	GGCCCCCCC
2401	AGCGCCGCCA	GCGCCTTCGT	GGGCGCCGGC	ATCGCCGGCG	CCGCCGTGGG
2451	CAGCATCGGC	CTGGGCAAGG	TGCTGGTGGA	CATCCTGGCC	GGCTACGGCG
2501	CCGGCGTGGC	CGGCGCCCTG	GTGGCCTTCA	AGGTGATGAG	CGGCGAGATG
2551	CCCAGCACCG	AGGACCTGGT	GAACCTGCTG	CCCGCCATCC	TGAGCCCCGG
2601	CGCCCTGGTG	GTGGGCGTGG	TGTGCGCCGC	CATCCTGCGC	CGCCACGTGG
2651	GCCCCGGCGA	GGGCGCCGTG	CAGTGGATGA	ACCGCCTGAT	CGCCTTCGCC
2701	AGCCGCGGCA	ACCACGTGAG	CCCCACCCAC	TACGTGCCCG	AGAGCGACGC
2751	CGCCGCCCGC	GTGACCCAGA	TCCTGAGCAG	CCTGACCATC	ACCCAGCTGC
2801	TGAAGCGCCT	GCACCAGTGG	ATCAACGAGG	ACTGCAGCAC	CCCCTGCAGC
2851	GGCAGCTGGC	TGCGCGACGT	GTGGGACTGG	ATCTGCACCG	TGCTGACCGA
2901	CTTCAAGACC	TGGCTGCAGA	GCAAGCTGCT	GCCCCAGCTG	CCCGGCGTGC
2951	CCTTCTTCAG	CTGCCAGCGC	GGCTACAAGG	GCGTGTGGCG	CGGCGACGGC
3001	ATCATGCAGA	CCACCTGCCC	CTGCGGCGCC	CAGATCACCG	GCCACGTGAA
3051	GAACGGCAGC	ATGCGCATCG	TGGGCCCCAA	GACCTGCAGC	AACACCTGGC
3101	ACGGCACCTT	CCCCATCAAC	GCCTACACCA	CCGGCCCCTG	CACCCCCAGC
3151	CCCGCCCCCA	ACTACAGCCG	CGCCCTGTGG	CGCGTGGCCG	CCGAGGAGTA
3201	CGTGGAGGTG	ACCCGCGTGG	GCGACTTCCA	CTACGTGACC	GGCATGACCA
3251	CCGACAACGT	GAAGTGCCCC	TGCCAGGTGC	CCGCCCCGA	GTTCTTCACC
3301	GAGGTGGACG	GCGTGCGCCT	GCACCGCTAC	GCCCCGCCT	GCCGCCCCT
3351	GCTGCGCGAG	GAGGTGACCT	TCCAGGTGGG	CCTGAACCAG	TACCTGGTGG

3401	CCACCCACCT	CCCCTCCGAG	CCCGAGCCCG	ACGTGGCCGT	GCTGACCAGC
3451			CATCACCGCC		
3501					AGCCAGCTGA
3551			ACCTGCACCA		
3601			CCTGCTGTGG		
3651			AGAACAAGGT		
			GACGAGCGCG		
3701			GACGAGCGCG		
3751					
3801			TGCTGGAGAG		
3851			TGCCCCCTGC		
3901			GCGCACCGTG		
3951	-		TGGCCACCAA		
4001			ACCGCCACCG		
4051			CGACGTGGAG		
4101			ACCCCGACCT		
4151	CCGTGAGCGA	GGAGGCCAGC	GAGGACGTGG	TGTGCTGCAG	CATGAGCTAC
4201	ACCTGGACCG	GCGCCCTGAT	CACCCCTGC	GCCGCCGAGG	AGAGCAAGCT
4251	GCCCATCAAC	GCCCTGAGCA	ACAGCCTGCT	GCGCCACCAC	AACATGGTGT
4301	ACGCCACCAC	CAGCCGCAGC	GCCGGCCTGC	GCCAGAAGAA	GGTGACCTTC
4351	GACCGCCTGC	AGGTGCTGGA	CGACCACTAC	CGCGACGTGC	TGAAGGAGAT
4401	GAAGGCCAAG	GCCAGCACCG	TGAAGGCCAA	GCTGCTGAGC	GTGGAGGAGG
4451	CCTGCAAGCT	GACCCCCCC	CACAGCGCCA	AGAGCAAGTT	CGGCTACGGC
4501	GCCAAGGACG	TGCGCAACCT	GAGCAGCAAG	GCCGTGAACC	ACATCCACAG
4551	CGTGTGGAAG	GACCTGCTGG	AGGACACCGT	GACCCCCATC	GACACCACCA
4601	TCATGGCCAA	GAACGAGGTG	TTCTGCGTGC	AGCCCGAGAA	GGGCGGCCGC
4651	AAGCCCGCCC	GCCTGATCGT	GTTCCCCGAC	CTGGGCGTGC	GCGTGTGCGA
4701	GAAGATGGCC	CTGTACGACG	TGGTGAGCAC	CCTGCCCCAG	GTGGTGATGG
4751	GCAGCAGCTA	CGGCTTCCAG	TACAGCCCCG	GCCAGCGCGT	GGAGTTCCTG
4801	GTGAACACCT	GGAAGAGCAA	GAAGAACCCC	ATGGGCTTCA	GCTACGACAC
4851	CCGCTGCTTC	GACAGCACCG	TGACCGAGAA	CGACATCCGC	GTGGAGGAGA
4901	GCATCTACCA	GTGCTGCGAC	CTGGCCCCCG	AGGCCCGCCA	GGCCATCAAG
4951			CATCGGCGGC		
5001	CCAGAACTGC	GGCTACCGCC	GCTGCCGCGC	CAGCGGCGTG	CTGACCACCA
5051	GCTGCGGCAA	CACCCTGACC	TGCTACCTGA	AGGCCAGCGC	CGCCTGCCGC

5101	GCCGCCAAGC	TGCAGGACTG	CACCATGCTG	GTGAACGCCG	CCGGCCTGGT
5151	GGTGATCTGC	GAGAGCGCCG	GCACCCAGGA	GGACGCCGCC	AGCCTGCGCG
5201	TGTTCACCGA	GGCCATGACC	CGCTACAGCG	CCCCCCCGG	CGACCCCCC
5251	CAGCCCGAGT	ACGACCTGGA	GCTGATCACC	AGCTGCAGCA	GCAACGTGAG
5301	CGTGGCCCAC	GACGCCAGCG	GCAAGCGCGT	GTACTACCTG	ACCCGCGACC
5351	CCACCACCCC	CCTGGCCCGC	GCCGCCTGGG	AGACCGCCCG	CCACACCCCC
5401	GTGAACAGCT	GGCTGGGCAA	CATCATCATG	TACGCCCCCA	CCCTGTGGGC
5451	CCGCATGATC	CTGATGACCC	ACTTCTTCAG	CATCCTGCTG	GCCCAGGAGC
5501	AGCTGGAGAA	GGCCCTGGAC	TGCCAGATCT	ACGGCGCCTG	CTACAGCATC
5551	GAGCCCCTGG	ACCTGCCCCA	GATCATCGAG	CGCCTGCACG	GCCTGAGCGC
5601	CTTCAGCCTG	CACAGCTACA	GCCCCGGCGA	GATCAACCGC	GTGGCCAGCT
5651	GCCTGCGCAA	GCTGGGCGTG	CCCCCCTGC	GCGTGTGGCG	CCACCGCGCC
5701	CGCAGCGTGC	GCGCCCGCCT	GCTGAGCCAG	GGCGGCCGCG	CCGCCACCTG
5751	CGGCAAGTAC	CTGTTCAACT	GGGCCGTGAA	GACCAAGCTG	AAGCTGACCC
5801	CCATCCCCGC	CGCCAGCCAG	CTGGACCTGA	GCGGCTGGTT	CGTGGCCGGC
5851	TACAGCGGCG	GCGACATCTA	CCACAGCCTG	AGCCGCGCCC	GCCCCCGCTG
5901	GTTCATGCTG	TGCCTGCTGC	TGCTGAGCGT	GGGCGTGGGC	ATCTACCTGC
5951	TGCCCAACCG	CTAAA			

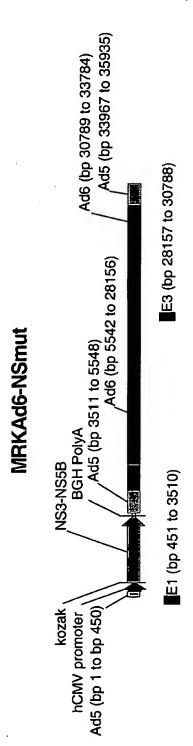


FIG. 4A

1	catcatcast	aatatacctt	attttggatt	gaagccaata	tgataatgag	ggggtggagt
7	thetesesta	acacacacac	taraacaa	gcgggtgacg	tagtagtgtg	gcggaagtgt
121	-stattasss	gtgtgggggg	acacatotaa	gcgacggatg	togcaaaagt	gacgtttttg
101	gatgttgtaa	gtgtggegga	deactracaa	ttttcgcgcg	attttaggcg	gatgttgtag
101	tabatttaga	cataaccaa	taagatttag	ccattttcgc	gggaaaactg	aataagagga
241	cadacttygg	rastasttt	atattactca	tagcgcgtaa	tatttgtcta	agaccacaga
301	agtgaaattt	atttacataa	anactonco	aggtgttttt	ctcaggtgtt	ttccqcqttc
701 701	gactitgacc	ttaacatttt	attattatan	gcggccgcga	tccattgcat	acattatate
401	cygytcadag	tatatacatt	tatattooct	catgtccaac	attaccocca	tottgacatt
401	catattataa	tagttatta	tacteatcaa	ttacggggtc	attagttcat	agcccatata
601	tagagettaga	cattacataa	cttacaataa	atggcccgcc	togctgaccg	cccaacgacc
00I	tggagtteeg	cgccacacaa	atracratato	ttcccatagt	aacgccaata	gggactttcc
221	attacates	ataaataaa	tatttaccot	aaactgccca	cttggcagta	catcaagtgt
721	attgatgtta	acgggcggag	cctattgacq	tcaatgacgg	taaatggccc	gcctggcatt
041	attacacacata	catcacctta	taggactttc	ctacttggca	gtacatctac	gtattagtca
041	togotattac	catgateta	caattttaac	agtacatcaa	tagacataga	tagcggtttg
901	regulaciac	atttccaact	ctccacccca	ttgacgtcaa	taggagttta	ttttggcacc
1021	actcacgggg	ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg
1021	addatedacy	agactecea	atctatataa	gcagagctcg	tttagtgaac	catcagatca
1141	graggrage	acggcgggag	tattttaacc	tccatagaag	acaccoooac	cgatccagcc
1141	tereserres	gazagatas	attogaaco	ggattccccg	taccaagagt	gagatetgee
1201	ceegeggeeg	ggaacggtgt	ctactcccaa	cagacgcggg	acctacttag	ttgcatcatc
1201	accatggege	ccaccacggc	Caacaaccad	gtcgagggag	aggttcaggt	ggtttccacc
1321	accageerra	catteetee	caagaaccag	aacggcgtgt	attagaccat	ttaccatggt
1301	gcaacacaac	agacettage	caacccaaaa	gggccaatca	cccagatgta	cactaatqtq
1441	getggeteaa	togtoggeta	acadacacac	cccggggcgc	gttccttgac	accatocacc
1201	gaccaggacc	cogcoggoog	cttaatcaca	agacatgctg	acotcattco	aatacaccaa
1001	tgtggcagct	cagaccerea	cctactctcc	cccaggcctg	tetectactt	gaagggctct
1021	eggggegaca	gragggggag	cccttcaaa	cacgctgtgg	gcatcttccg	ggctgccgta
1001	tegggtggte	gagttacass	agcaataac	tttgtgcccg	tagagtccat	ggaaactact
1/41	tgcacccggg	gggttgtgaa	ggeggeggae	tccccccgg	ccgtaccgca	gtcatttcaa
1001	atgeggtete	tacacactec	cactoocaoc	ggcaagagta	ctaaagtgcc	ggctgcatat
1001	geggeeeace	gatacaagat	acteateate	aatccgtccg	ttgccgctac	cttagggttt
1921	geagectaag	tatataagga	acacqutatt	gaccccaaca	tcagaactgg	ggtaaggacc
1301	ggggcgtata	acacacacat	cacatactct	acctatggca	agtttcttgc	cgatggtggt
2041	tactactacag	gegeeeege	catcataata	tgtgatgagt	gccattcaac	tgactcgact
210I	cgccccgggg	gcgcccacga	eatcataata	caagcggaga	caactagaac	acaacttate
2221	ataattttgg	contact	tecaggae	gtcaccgtgc	cacacccaaa	catcgaggag
2221	graceratat	ctaatactcc	agagatoco	ttctatggca	aagccatccc	cattgaagcc
2201	gtggccccgc	gaaggeatet	cattttctgt	cattccaaga	agaagtgcga	cgagetegee
234T	accagggggg	carreter	aatcaacoct	gtggcgtatt	accagagact	cgatgtgtcc
2401	gtaaagttgt	ctatcccaga	catcattate	gtggcaacag	acgetetgat	gacgggctat
2401	gccacaccaa	ttaactcagt	datcdactdt	aacacatgtg	tcacccagac	agtcgacttc
2221	acgygegact	coacttosc	cattgagagaga	acgaccgtgc	ctcaagacgc	agtgtcgcgc
228T	agettggate		taccagaca	aggagaggca	tctacaggtt	tataactcca
2641	tegeagegge	ggggtaggat	atteasttee	tcggtcctgt	gtgagtgcta	tgacgcgggc
2701	ggagaacggc	cereggeat	geeggeegg	acctcggtta	gagagagaa	ctacctgaac
2761	tgtgcttggt	tacagattta	ccecgecgag	ctggagttct	aggagagtat	cttcacaggc
5851	acaccagggu	tgeeegeeg	ctaggactac	cagaccaagc	addcaddada	caacttcccc
7881	ctcacccaca	cayatycaca	cacactatas	gccagggctc	-aaccccacc	tccatcatoo
2941	racciggiag	ggaagtgtgt	catacaacta	aaacctacgc	tacacagace	aacaccctto
300T	yatcaaatgt	tagazgaset	ccassators	gtcaccctca	cccaccccat	aaccaaatac
3061	ctgtacaggc	agatataaa	tancetana	gtcgtcacta	acacctagat	actaataaac
3121	accatggcat	geargregge	cacatattac	ctgacaacag	acaatataat	cattotogot
2181	ggagreettg	tatacasas	acceptate	gttcccgaca	addagtttct	ctaccaggag
3241	aggattatet	Lgtccgggag	googgetatt	gillingala	gggageeee	

3361	gccgagcaat	tcaagcagaa	agcgctcggg	ttactgcaaa	cagccaccaa	acaagcggag
3421	gctgctgctc	ccgtggtgga	gtccaagtgg	cgagcccttg	agacattctg	ggcgaagcac
3481	atgtggaatt	tcatcagcgg	gatacagtac	ttagcaggct	tatccactct	gcctgggaac
3541	cccgcaatag	catcattgat	ggcattcaca	gcctctatca	ccagcccgct	caccacccaa
3601	agtaccetee	tgtttaacat	cttggggggg	tgggtggctg	cccaactcgc	ccccccagc
3661	gccgcttcgg	ctttcgtggg	cgccggcatc	gccggtgcgg	ctgttggcag	cataggcctt
3721	gggaaggtgc	ttgtggacat	tctggcgggt	tatggagcag	gagtggccgg	cgcgctcgtg
3781	gccttcaagg	tcatgagcgg	cgagatgccc	tccaccgagg	acctggtcaa	tctacttcct
3841	gccatcctct	ctcctggcgc	cctggtcgtc	ggggtcgtgt	gtgcagcaat	actgcgtcga
3901	cacgtgggtc	cgggagaggg	ggctgtgcag	tggatgaacc	ggctgatagc	gttcgcctcg
3961	cggggtaatc	atgtttcccc	cacgcactat	gtgcctgaga	gcgacgccgc	agcgcgtgtt
	actcagatcc					
	aatgaagact					
4141	tgcacggtgt	tgactgactt	caagacctgg	ctccagtcca	agctcctgcc	gcagctaccg
4201	ggagtccctt	ttttctcgtg	ccaacgcggg	tacaagggag	tctggcgggg	agacggcatc
	atgcaaacca					
	aggatcgtcg					
	tacaccacgg					
4441	gtggccgctg	aggagtacgt	ggaggtcacg	cgggtggggg	atttccacta	cgtgacgggc
4501	atgaccactg	acaacgtaaa	gtgcccatgc	caggttccgg	ctcctgaatt	cttcacggag
4561	gtggacggag	tgcggttgca	caggtacgct	ccggcgtgca	ggcctctcct	acgggaggag
4621	gttacattcc	aggtcgggct	caaccaatac	ctggttgggt	cacagctacc	atgcgagccc
4681	gaaccggatg	tagcagtgct	cacttccatg	ctcaccgacc	cctcccacat	cacagcagaa
4741	acggctaagc	gtaggttggc	cagggggtct	ccccctcct	tggccagctc	ttcagctagc
4801	cagttgtctg	cgccttcctt	gaaggcgaca	tgcactaccc	accatgtctc	tccggacgct
	gacctcatcg					
	gagtcggaga					
4981	gagagggaag	tatccgttcc	ggcggagatc	ctgcggaaat	ccaagaagtt	ccccgcagcg
5041	atgcccatct	gggcgcgccc	ggattacaac	cctccactgt	tagagtcctg	gaaggacccg
5101	gactacgtcc	ctccggtggt	gcacgggtgc	ccgttgccac	ctatcaaggc	ccctccaata
5161	ccacctccac	ggagaaagag	gacggttgtc	ctaacagagt	cctccgtgtc	ttctgcctta
5221	gcggagctcg	ctactaagac	cttcggcagc	tccgaatcat	cggccgtcga	cagcggcacg
	gcgaccgccc					
	tactcctcca					
5401	tggtctaccg	tgagcgagga	agctagtgag	gatgtcgtct	gctgctcaat	gtcctacaca
	tggacaggcg					
5521	ttgagcaact	ctttgctgcg	ccaccataac	atggtttatg	ccacaacatc	tcgcagcgca
	ggcctgcggc					
	gacgtgctca					
	gaggaagcct					
5761	aaggacgtcc	ggaacctatc	cagcaaggcc	gttaaccaca	tccactccgt	gtggaaggac
	ttgctggaag					
5881	tgtgtccaac	cagagaaagg	aggccgtaag	ccagcccgcc	ttatcgtatt	cccagatctg
	ggagtccgtg					
	gtgatgggct					
6061	aatacctgga	aatcaaagaa	aaaccccatg	ggcttttcat	atgacactcg	ctgtttcgac
	tcaacggtca					
	gccccgaag					
	ctgactaatt					
	acgactagct					
	gcgaagctcc					
6421	agcgcgggaa	cccaagagga	cgcggcgagc	ctacgagtct	tcacggaggc	tatgactagg
	tactctgccc					
	tgttcctcca					
6601	cgtgatccca	ccacccccct	cgcacgggct	gcgtgggaaa	cagctagaca	cactccagtt

6661	aactcctggc	taggcaacat	tatcatgtat	gcgcccactt	tgtgggcaag	gatgattctg
6721	atgactcact	tcttctccat	ccttctagca	caggagcaac	ttgaaaaagc	cctggactgc
6781	cagatctacg	gggcctgtta	ctccattgag	ccacttgacc	tacctcagat	cattgaacga
6841	ctccatggcc	ttagcgcatt	ttcactccat	agttactctc	caggtgagat	caatagggtg
6901	gcttcatgcc	tcaggaaact	tggggtacca	cccttgcgag	tctggagaca	tegggeeagg
6961	agcatccaca	ctaggctact	gtcccagggg	gggagggccg	ccacttgtgg	caagtacctc
7021	ttcaactggg	cagtgaagac	caaactcaaa	ctcactccaa	tcccggctgc	gtcccagctg
7081	gacttgtccg	actaattcat	tgctggttac	agcgggggag	acatatatca	cagectgtet
7141	cotocccac	cccactaatt	catgctgtgc	ctactcctac	tttctgtagg	ggtaggcate
7201	tacctoctcc	ccaaccggta	aatctagagc	tgtgccttct	agttgccagc	catctgttgt
7261	ttacccctcc	cccatacctt	ccttgaccct	ggaaggtgcc	actcccactg	tectttecta
7321	ataaaatgag	gaaattgcat	cgcattgtct	gagtaggtgt	cattctattc	tggggggtgg
7381	aataaaacaa	gacagcaagg	gggaggattg	ggaagacaat	agcaggcatg	ctggggatgc
7441	ggtgggctct	atggccgatc	ggcgcgccgt	actgaaatgt	gtgggcgtgg	cttaagggtg
7501	ggaaagaata	tataaggtgg	gggtcttatg	tagttttgta	tctgttttgc	agcagccgcc
7561	accaccataa	gcaccaactc	gtttgatgga	agcattgtga	gctcatattt	gacaacgcgc
7621	atgcccccat	gggccggggt	gcgtcagaat	gtgatgggct	ccagcattga	tggtcgcccc
7681	atactacca	caaactctac	taccttgacc	tacgagaccg	tgtctggaac	gccgttggag
7741	actocaocct	ccaccaccac	ttcagccgct	gcagccaccg	cccgcgggat	tgtgactgac
7801	tttactttcc	tgagcccgct	tgcaagcagt	gcagcttccc	gttcatccgc	ccgcgatgac
7861	aagttgacgg	ctcttttqqc	acaattggat	tctttgaccc	gggaacttaa	tgtegtttet
7921	cagcagctgt	tggatctgcg	ccagcaggtt	tctgccctga	aggetteete	cccccaat
7981	gcggtttaaa	acataaataa	aaaaccagac	tctgtttgga	tttggatcaa	gcaagtgtet
8041	tgctgtcttt	atttaggggt	tttgcgcgcg	cggtaggccc	gggaccagcg	gtctcggtcg
8101	ttgagggtcc	tgtgtatttt	ttccaggacg	tggtaaaggt	gactctggat	gttcagatac
8161	atgggcataa	gcccgtctct	ggggtggagg	tagcaccact	gcagagcttc	atgetgeggg
8221	gtggtgttgt	agatgatcca	gtcgtagcag	gagcgctggg	cgtggtgcct	aaaaatgtct
8281	ttcagtagca	agctgattgc	caggggcagg	cccttggtgt	aagtgtttac	aaagcggtta
8341	agctgggatg	ggtgcatacg	tggggatatg	agatgcatct	tggactgtat	ttttaggttg
8401	gctatgttcc	cagccatatc	cctccgggga	ttcatgttgt	gcagaaccac	cagcacagtg
8461	tatccggtgc	acttgggaaa	tttgtcatgt	agcttagaag	gaaatgcgtg	gaagaacttg
8521	gagacgccct	tgtgacctcc	aagattttcc	atgcattcgt	ccataatgat	ggcaatgggc
8581	ccacgggcgg	cggcctgggc	gaagatattt	ctgggatcac	taacgtcata	gttgtgttcc
8641	aggatgagat	cgtcataggc	catttttaca	aagcgcgggc	ggagggtgcc	agactgcggt
8701	ataatggttc	catccggccc	aggggcgtag	ttaccctcac	agatttgcat	ttcccacyct
8761	ttgagttcag	atggggggat	catgtctacc	tgcggggcga	tgaagaaaac	ggtttccggg
8821	gtaggggaga	tcagctggga	agaaagcagg	ttcctgagca	gctgcgactt	accycayccy
8881	gtgggcccgt	aaatcacacc	tattaccggc	tgcaactggt	agttaagaga	getgeagetg
8941	ccgtcatccc	tgagcagggg	ggccacttcg	ttaagcatgt	ccctgactcg	catgttttcc
9001	ctgaccaaat	ccgccagaag	gcgctcgccg	cccagcgata	gcagttcttg	Caaggaagca
9061	aagtttttca	acggtttgag	accgtccgcc	gtaggcatgc	ttttgagcgt	ccgaccaage
9121	agttccaggc	ggtcccacag	ctcggtcacc	tgctctacgg	catctcgatc	tagtagagagag
9181	cctcgtttcg	cgggttgggg	cggctttcgc	tgtacggcag	tagtcggtgc	tegtecayae
9241	gggccagggt	catgtctttc	cacgggcgca	gggtcctcgt	cagcgtagtc	tgggtcatgg
9301	tgaaggggtg	cgctccgggc	tgcgcgctgg	ccagggtgcg	cttgaggetg	gtcctgctgg
9361	tgctgaagcg	ctgccggtct	tegeeetgeg	cgtcggccag	gtagcatttg	accatggtgt
9421	catagtccag	cccctccgcg	gcgtggccct	tggcgcgcag	ettgecettg	gaggaggcgc
9481	cgcacgaggg	gcagtgcaga	cttttgaggg	cgtagagctt	gggcgcgaga	adiaccyatt
9541	ccggggagta	ggcatccgcg	ccgcaggccc	cgcagacggt	ctcgcattcc	acgagecagg
9601	tgagctctgg	ccgttcgggg	tcaaaaacca	ggtttccccc	atgetttttg	atgcgtttct
9661	tacctctggt	ttccatgagc	cggtgtccac	gctcggtgac	gaaaaggctg	teegtgteee
9721	cgtatacaga	cttgagaggc	ctgtcctcga	gcggtgttcc	geggteetee	tcgtatagaa
9781	actcggacca	ctctgagacg	aaggctcgcg	tccaggccag	cacgaaggag	gctaagtggg
9841	aggggtagcg	gtcgttgtcc	actagggggt	ccactcgctc	cagggtgtga	agacacatgt
9901	cgccctcttc	ggcatcaagg	aaggtgattg	gtttataggt	graggecacg	tgaccgggtg

9961	ttcctgaagg	ggggctataa	aagggggtgg	gggcgcgttc	gtcctcactc	tcttccgcat
		gagggccagc				
		attgtcagtt				
						ttgttgtcaa
		aaacgacccg				
10261	tttggttttt	gtcgcgatcg	gcgcgctcct	tggccgcgat	gtttagctgc	acgtattcgc
		ccgccattcg				
10381	gccaaccgcg	gttgtgcagg	gtgacaaggt	caacgctggt	ggctacctct	ccgcgtaggc
10441	gctcgttggt	ccagcagagg	cggccgccct	tgcgcgagca	gaatggcggt	agtgggtcta
10501	gctgcgtctc	gtccgggggg	tctgcgtcca	cggtaaagac	cccgggcagc	aggcgcgcgt
		tatcttgcat				
		gtatgggttg				
		gcaaatgtcg				
10741	ggtagcatct	tccaccgcgg	atgctggcgc	gcacgtaatc	gtatagttcg	tgcgagggag
		gggaccgagg				
		atgtgagttg				
		taccgcgtca				
10981	gctcggcggt	gacctgcacg	tctagggcgc	agtagtccag	ggtttccttg	atgatgtcat
11041	acttatcctg	tcccttttt	ttccacagct	cgcggttgag	gacaaactct	tcgcggtctt
11101	tccagtactc	ttggatcgga	aacccgtcgg	cctccgaacg	gtaagagcct	agcatgtaga
		ggcctggtag				
		gagcgaggtg				
		gaagtcagtg				
11341	gctttttgga	acgcgggttt	ggcagggcga	aggtgacatc	gttgaagagt	atctttcccg
11401	cgcgaggcat	aaagttgcgt	gtgatgcgga	agggtcccgg	cacctcggaa	cggttgttaa
11461	ttacctgggc	ggcgagcacg	atctcgtcaa	agccgttgat	gttgtggccc	acaatgtaaa
11521	gttccaagaa	gcgcgggatg	cccttgatgg	aaggcaattt	tttaagttcc	tcgtaggtga
11581	gctcttcagg	ggagctgagc	ccgtgctctg	aaagggccca	gtctgcaaga	tgagggttgg
11641	aagcgacgaa	tgagctccac	aggtcacggg	ccattagcat	ttgcaggtgg	tcgcgaaagg
11701	tcctaaactg	gcgacctatg	gccattttt	ctggggtgat	gcagtagaag	gtaagcgggt
11761	cttgttccca	gcggtcccat	ccaaggtccg	cggctaggtc	tcgcgcggcg	gtcactagag
		gccgaacttc				
11881	ccatccaagt	ataggtctct	acatcgtagg	tgacaaagag	acgctcggtg	cgaggatgcg
11941	agccgatcgg	gaagaactgg	atctcccgcc	accagttgga	ggagtggctg	ttgatgtggt
12001	gaaagtagaa	gtccctgcga	cgggccgaac	actcgtgctg	gcttttgtaa	aaacgtgcgc
12061	agtactggca	gcggtgcacg	ggctgtacat	cctgcacgag	gttgacctga	cgaccgcgca
12121	caaggaagca	gagtgggaat	ttgagcccct	cgcctggcgg	gtttggctgg	tggtcttcta
12181	cttcggctgc	ttgtccttga	ccgtctggct	gctcgagggg	agttacggtg	gatcggacca
12241	ccacgccgcg	cgagcccaaa	gtccagatgt	ccgcgcgcgg	cggtcggagc	ttgatgacaa
12301	catcgcgcag	atgggagctg	tccatggtct	ggagctcccg	cggcgtcagg	tcaggcggga
		gtttacctcg				
12421	tgatttccag	gggctggttg	gtggcggcgt	cgatggcttg	caagaggccg	catccccgcg
12481	gcgcgactac	ggtaccgcgc	ggcgggcggt	gggccgcggg	ggtgtccttg	gatgatgcat
12541	ctaaaagcgg	tgacgcgggc	gggcccccgg	aggtaggggg	ggctcgggac	ccgccgggag
12601	agggggcagg	ggcacgtcgg	cgccgcgcgc	gggcaggagc	tggtgctgcg	cgcggaggtt
12661	gctggcgaac	gcgacgacgc	ggcggttgat	ctcctgaatc	tggcgcctct	gcgtgaagac
		gtgagcttga				
		tggcgcaaaa				
		tgctcgatct				
		tcgttggaga				
		cggctgtaga				
13021	cgcgagattg	agctccacgt	gccgggcgaa	gacggcgtag	tttcgcaggc	gctgaaagag
13081	gtagttgagg	gtggtggcgg	tgtgttctgc	cacgaagaag	tacataaccc	agcgccgcaa
13141	cgtggattcg	ttgatatccc	ccaaggcctc	aaggcgctcc	atggcctcgt	agaagtccac
13201	ggcgaagttg	aaaaactggg	agttgcgcgc	cgacacggtt	aactcctcct	ccagaagacg

13261	gatgageteg	gcgacagtgt	cgcgcacctc	gcgctcaaag	gctacagggg	cctcttcttc
13321	ttcttcaatc	tcctcttcca	taagggcctc	cccttcttct	tcttctggcg	gcggtggggg
13381	aggggggaca	cggcggcgac	gacggcgcac	cgggaggcgg	tcgacaaagc	gctcgatcat
13441	ctccccacaa	cgacggcgca	tggtctcggt	gacggcgcgg	ccgttctcgc	gggggcgcag
13501	ttggaagacg	ccaccatca	tgtcccggtt	atgggttggc	ggggggctgc	cgtgcggcag
13561	ggatacggcg	ctaacgatgc	atctcaacaa	ttgttgtgta	ggtactccgc	caccgaggga
13621	cctgagggag	tccgcatcga	ccggatcgga	aaacctctcg	agaaaggcgt	ctaaccagtc
13681	acagtcgcaa	ggtaggctga	gcaccgtggc	gggcggcagc	gggcggcggt	cggggttgtt
13741	tctggcggag	gtgctgctga	tgatgtaatt	aaagtaggcg	gtcttgagac	ggcggatggt
13801	cdacadaadc	accatotcct	taggtccggc	ctgctgaatg	cgcaggcggt	cggccatgcc
13861	ccaggetteg	ttttgacatc	ggcgcaggtc	tttgtagtag	tcttgcatga	gcctttctac
13921	concacttet	tcttctcctt	cctcttgtcc	tgcatctctt	gcatctatcg	ctgcggcggc
13981	ggcggagttt	ggccgtaggt	agegeeetet	tcctcccatg	cgtgtgaccc	cgaagcccct
14041	catcooctoa	agcagggcca	gatcaacaac	aacgcgctcg	gctaatatgg	cctgctgcac
14101	ctacataaga	gtagactgga	agtcgtccat	gtccacaaag	cggtggtatg	cgcccgtgtt
14161	gatggtgtaa	gtgcagttgg	ccataacgga	ccagttaacg	gtctggtgac	ccggctgcga
14221	gacggcgcaa	tacctgagac	acaagtaagc	ccttgagtca	aagacgtagt	cgttgcaagt
14221	ccacaccaaa	tactogtatc	ccaccaaaaa	atacaacaac	ggctggcggt	agaggggcca
14201	acatacaata	accagaactc	caaaaacaaa	gtcttccaac	ataaggcgat	gatatccgta
14341	gegeagggeg	gacatccagg	tgatgccggc	aacaataata	gaggcgcgcg	gaaagtcacg
14461	gacgcacttc	cagatgttgc	gcagcggcaa	aaagtgctcc	atggtcggga	cgctctggcc
14501	antcannon	acacagteat	tgacgctcta	gaccgtgcaa	aaggagagcc	tgtaagcggg
14521	cactetteed	taatctaata	gataaattcg	caagggtatc	atggcggacg	accggggttc
14501	descende	tecaaccate	caccataata	catgcggtta	ccgccgcgt	gtcgaaccca
14701	gatatacasc	gtcagacaac	adadaaacac	tecttttage	ttccttccag	gcgcggcgga
14761	tactacacta	actttttaa	ccactoocco	cacacaacat	aagcggttag	gctggaaagc
14901	gaaagcatta	agtagetege	tecetatage	cggagggtta	ttttccaagg	gttgagtcgc
14021	gaaagcacca	attcaagtct	caaaccaacc	ggactgcggc	gaacgggggt	ttgcctcccc
14001	gggacccccg	accccactta	caaattcctc	cogaaacagg	gacgagcccc	ttttttgctt
15001	ttcccacato	catcoggtag	tgcggcagat	acaccccct	cctcagcagc	ggcaagagca
15061	adadcadcad	cadacataca	gggggcgtc	cccttctcct	accgcgtcag	gaggggcaac
15121	atccccccct	dacacaacaa	cagatogtga	ttacgaaccc	ccgcggcgcc	ggacccggca
15101	ctacttogac	ttagaggagg	acaaaaacct	agcacaacta	ggagcgccct	ctcctgagcg
15101	acacccaacc	atacaactaa	agcgtgacac	acacasaaca	tacgtgccgc	ggcagaacct
15241	atttcccaagg	gegeageega	aggaggggaa	ggagatgcgg	gatcgaaagt	tccatgcagg
15361	gccccgcgac	caccataacc	tgaaccgcga	acaattacta	cgcgaggagg	actttgagcc
15301	gegegageeg	accognatta	atcccacaca	cacacacata	gcggccgccg	acctggtaac
15421	cgacgcgcgg	cacaccottoa	accaggagat	taactttcaa	aaaagcttta	acaaccacgt
10401	acacacactt	ataacacaca	aggaggtggc	tataggactg	atgcatctgt	gggactttgt
12241	agegeacycer	gagagagag	caaatagcaa	accactcata	gcgcagctgt	tccttatagt
12001	aagcycycty	addegagaga	aggrattcag	ggatgcgctg	ctaaacatag	tagagcccga
15721	geageacage	ctactcaatt	tgataaacat	tctgcagagc	atagtggtgc	aggagcgcag
15701	ottoageeta	actascasa	tggccgccat	taactattcc	atgctcagtc	tgggcaagtt
15/01	ttagageeeg	acceptatace	atacccctta	cottcccata	gacaaggagg	taaagatcga
15001	accettates	atgatatace	cactasagat	gettacetta	agcgacgacc	tgggcgttta
15901	bassassass	castcas	aggegaagge	cataaaccaa	cggcgcgagc	tcagcgaccg
12301	ceyeaacyay	cgcacccaca	aggeegegag	anctageaga	ggcagcggcg	atagagaggc
10021	cgagetgatg	tttaacacaa	acactascet	acactagace	ccaagccgac	acaccctaga
16081	egagteetae	accepte	gegeegaeee	gegeegggee	cgcgctggca	acatcaacaa
16141	ggcagctggg	tatasasas	accatcacta	Cusuccausa	gacggcgagt	actaagcggt
16201	cgtggaggaa	tatgacgagg	acyatyayta acaacaca	accoraccora	cggtgcgggc	aacactacaa
10201	gatgittetg	accayacyac	ctccacacac	dactddcacc	aggtcatgga	ccgcatcatg
16321	agccagccgt	eeggeettaa	taacacatta	Cuucsucsuc	cgcaggccaa	ccaactctcc
16381	ccgccgaccg	cgcgcaaccc	coargegee	Transcrer	cgcacgagaa	agtactage
16441	gcaattctgg	aagcygtggt	22292929292	ateccaceca	atgaggccgg	cctaatctac
16501	accgtaaacg	cgctggccga	aaacagggcc	accoggeecy	argaggeegg	

1						
10201	gacgcgctgc	ttcagcgcgt	ggctcgttac	aacagcagca	acgtgcagac	caacetggac
					agcgcgcgca	
					cacagcccgc	
					ggctaatggt	
16801	ccgcaaagtg	aggtgtatca	gtccgggcca	gactattttt	tccagaccag	tagacaaggc
					aggggctgtg	
					cgcccaactc	
					cccgggacac	
					atgtggacga	
					acacgggcag	
					cctcgttgca	
17221	accetgaace	acceptigat	acactatata	cadcadadco	tgagccttaa	cctgatgcgc
17201	agegaggagg	agegeaceet	gagactageg	atraccrcrc	gcaacatgga	acconnecto
					acttgcatcg	
					actggctacc	
					gattcctctg	
					agttgcaaca	
					gcagcttgtc	
					gcttgatagg	
					agtacctaaa	
17761	ctgcagccgc	agcgcgaaaa	gaacctgcct	ccggcgtttc	ccaacaacgg	gatagagagc
17821	ctagtggaca	agatgagtag	atggaagacg	tatgcgcagg	agcacaggga	tgtgcccggc
17881	ccgcgcccgc	ccacccgtcg	tcaaaggcac	gaccgtcagc	ggggtctggt	gtgggaggac
					ggagtggcaa	
					catgatgcaa	
					ccccttagta	
					gtggtgagcg	
18181	adcaacaaca	ctagattcac	ccttcgatgc	teceetgase	ccgccgttcg	tacctccaca
10201	ggcggcggcg	cctaccaaa	dagagaga	catccottac	tctgagttgg	cacccctatt
					gatgtggcat	
					aacaatgact	
					cactggggcg	
					atgtttacca	
18541	ggcgcgggtg	atggtgtcgc	getegettae	taaggacaaa	caggtggagc	tgaaatacga
					accatgacca	
					cagaacgggg	
					gggtttgacc	
18781	tcttgtcatg	cctggggtat	atacaaacga	agccttccat	ccagacatca	ttttgctgcc
18841	aggatgcggg	gtggacttca	cccacagccg	cctgagcaac	ttgttgggca	tccgcaagcg
18901	gcaacccttc	caggagggct	ttaggatcac	ctacgatgac	ctggagggtg	gtaacattcc
18961	cgcactgttg	gatgtggacg	cctaccaggc	aagcttgaaa	gatgacaccg	aacagggcgg
					gaagagaact	
					gccattcgcg	
19141	toccacacoo	gcggaggaga	agcgcgctga	ggccgaggca	gcggccgaag	ctgccgcccc
					aagaaaccgg	
					agcaatgaca	
					cctcaggccg	
					gagcaggtat	
					cgccagatca	
					ttctacaacg	
					gtgttcaatc	
					accgtcagtg	
					atcggaggag	
					tacaaggccc	
19801	ctcgccgcgc	gtcctatcga	gccgcacttt	ttgagcaagc	atgtccatcc	ttatatcgcc

10061		acaggctggg	acctacactt	сссаапсаап	atatttagca	gggccaagaa
19001	cagcaacaac	caacacccag	tacacataca	coordinates	cacacaccct	ggggcgcgca
19921	gegetetgat	cgcactgggc	acaccaccat	cdatdacdcc	atcgacgcgg	tggtggagga
13301	caaacgcggc	tacacgccca	caccaccacc	agtgtcacc	gtggacgcgg	ccattcagac
20041	ggegegeaac	ggagcccggc	actacactaa	aatgaagaga	caacaaaaac	gcgtagcacg
20101	egregerees	cgccgacccg	gctacgctaa	ccaacacaca	acaacaaccc	tocttaacco
20101	tegecacege	accggccgac	gcaccaccat	acaaacaact	cgaaggctgg	ccacaaatat
20221	egeaegtege	cccccaggt	ccaaacaaca	accaccacc	acaacaacca	caaccattaa
20281	tgteactgtg	cagggtcgca	ccaggegacg	atactaata	cacaactcaa	ttagcggcct
20341	tgctatgact	gtgcgcaccc	ggggcaacgt	caactagggtg	gcaataaaaa	actacttaga
20401	gegegtgeee	tgtatgtatc	caccaccac	aacacagate	gaagetatgt	ccaagcgcaa
20461	ctcgtactgt	gagatgctcc	aggregation	accaraate	tatggccccc	cgaagaagga
20521	aatcaaagaa	tacaagcccc	dagaccatege	acaaatcaaa	aagaaaaaga	aagatgatga
2028T	agagcaggat	cttgacgacg	aggetaaa	attacacaca	accococcca	aacaacaaat
20641	tgatgatgaa	ggtcgacgcg	taacacctct	tttgcgaccc	aacaccacca	tagtctttac
20701	acagtggaaa	cgctccaccc	gcacctacaa	acacatatat	gatgaggtgt	acggcgacga
20/61	gcccggcgag	gagcaggcca	acceptacet	caaaaaattt	acctacagaa	agcggcataa
20821	ggacctgctt	gageaggeea	tegagegeee	caacccaaca	cctaccctaa	agcccgtgac
20881	ggacatgctg	gcgttgccgc	cocttocaco	atcccaaca	aaacacaacc	taaagcgcga
20941	actgcagcag	gtgctgcccg	cycligiaci	geecgaagaa	aagcgcggcc	gactggaaga
21001	gtctggtgac	ttggcaccca	ccgtgcaget	gatggtaccc	aagegeeage	tacaaccaat
21061	tgtcttggaa	aaaatgaccg	tggageetgg	gccggagccc	atteacatac	ccaccaccag
21121	caagcaggtg	gcaccgggac	egggegtgea	gaccgcggac	casacatoco	caattacctc
21181	tagcactagt	attgccactg	ccacagaggg	catggagaca	caaacgcccc	cctctaccca
21241	ggcggtggca	gatgccgcgg	tgcaggcggc	egetgeggee	gegeetaaga	accattcaaa
21301	ggtgcaaacg	gacccgtgga	tgtttcgtgt	tteageteec	ctacatectt	ccatcacaca
21361	gaagtacggc	gccgccagcg	egetactgee	cgaatatgee	cracaccccc	ccaccgcgcc
21421	tacccccggc	tatcgtggct	acacctaccg	ccccagaaga	cgagcaacca	cccgacgccg
21481	aaccaccact	ggaacccgcc	gccgccgtcg	cegtegeeag	ceegugeugg	cacactacca
21541	cgtgcgcagg	gtggctcgcg	aaggaggcag	gaecetggtg	Cigctaacag	teactacea
21601	ccccagcatc	gtttaaaagc	cggtctttgt	ggttcttgca	gatatggete	geatggeag
21661	cctccgtttc	ccggtgccgg	gattccgagg	aagaatgcac	cgtaggaggg	geatggeegg
21721	ccacggcctg	acgggcggca	tgcgtcgtgc	gcaccaccgg	eggeggegeg	egitegeactg
21781	tcgcatgcgc	ggcggtatcc	tgcccctcct	tattccactg	ategeegegg	cgattggtgt
21841	cgtgcccgga	attgcatccg	tggccttgca	ggcgcagaga	cactgattaa	aaacaaytta
21901	catgtggaaa	aatcaaaata	aaagtctgga	ctctcacgct	egettggtee	tgtaactatt
21961	ttgtagaatg	gaagacatca	actttgcgtc	actggccccg	cgacacggct	cgcgcccgtt
22021	catgggaaac	tggcaagata	tcggcaccag	caatatgagc	ggtggcgcct	teagetgggg
22081	ctcgctgtgg	agcggcatta	aaaatttcgg	ttccgccgtt	aagaactatg	geageaaage
22141	ctggaacagc	agcacaggcc	agatgctgag	ggacaagttg	aaagagcaaa	atttccaaca
22201	aaaggtggta	gatggcctgg	cctctggcat	tagcggggtg	gtggacctgg	ccaaccaggc
22261	agtgcaaaat	aagattaaca	gtaagcttga	tccccgccct	cccgtagagg	agcetecace
22321	aaccataaaa	acagtgtctc	cagaggggcg	tggcgaaaag	cgtccgcgac	ccgacaggga
22381	agaaactctg	gtgacgcaaa	tagacgagcc	tccctcgtac	gaggaggcac	taaagcaagg
22441	cctacccacc	acccgtccca	tcgcgcccat	ggctaccgga	gtgctgggcc	agcacacacc
22501	cataacacta	gacctgcctc	cccccgccga	cacccagcag	aaacctgtgc	tgccaggccc
22561	atccaccatt	gttgtaaccc	gtcctagccg	cgcgtccctg	cgccgcgccg	ccagcggtcc
22621	acaatcatta	caacccatag	ccagtggcaa	ctggcaaagc	acactgaaca	gcatcgtggg
22681	tttaaaaata	caatccctga	agcgccgacg	atgcttctga	tagctaacgt	gtcgtatgtg
22741	totcatotat	acatccatat	cgccgccaga	ggagctgctg	agccgccgcg	cgcccgcttt
22801	ccaagatggc	taccccttcg	atgatgccgc	agtggtctta	catgcacatc	tegggccagg
22861	acocctcoga	gtacctgage	cccgggctgg	tgcagttcgc	ccgcgccacc	gagacgtact
22921	tcagcctgaa	taacaagttt	agaaacccca	cggtggcgcc	tacgcacgac	gtgaccacag
22981	accogtetca	gcgtttgacg	ctgcggttca	tccccgtgga	ccgcgaggat	actgcgtact
23041	cotacaagge	acaattcacc	ctagctgtgg	gtgataaccg	tgtgctagac	atggetteea
23101	cgtactttga	catccacaac	gtgctggaca	ggggccctac	ttttaagccc	tactctggca
20101	-55					

23161	ctgcctacaa	cgcactggcc	cccaagggtg	cccccaactc	gtgcgagtgg	gaacaaaatg
23221	aaactgcaca	agtggatgct	caagaacttg	acgaagagga	gaatgaagcc	aatgaagctc
23281	aggcgcgaga	acaggaacaa	gctaagaaaa	cccatgtata	tgcccaggct	ccactgtccg
23341	gaataaaaat	aactaaagaa	ggtctacaaa	taggaactgc	cgacgccaca	gtagcaggtg
23401	ccggcaaaga	aattttcgca	gacaaaactt	ttcaacctga	accacaagta	ggagaatctc
23461	aatggaacga	agcggatgcc	acagcagctg	gtggaagggt	tcttaaaaag	acaactccca
23521	tgaaaccctg	ctatggctca	tacgctagac	ccaccaattc	caacggcgga	cagggcgtta
23581	tggttgaaca	aaatggtaaa	ttggaaagtc	aagtcgaaat	gcaattttt	tccacatcca
23641	caaatgccac	aaatgaagtt	aacaatatac	aaccaacagt	tgtattgtac	agcgaagatg
23701	taaacatgga	aactccagat	actcatcttt	cttataaacc	taaaatgggg	gataaaaatg
23761	ccaaagtcat	gcttggacaa	caagcaatgc	caaacagacc	aaattacatt	gcttttagag
23821	acaattttat	tggtctcatg	tattacaaca	gcacaggtaa	catgggtgtc	cttgctggtc
23881	aggcatcgca	gttgaacgct	gttgtagatt	tgcaagacag	aaacacagag	ctgtcctacc
23941	agcttttgct	tgattcaatt	ggcgacagaa	caagatactt	ttcaatgtgg	aatcaagctg
24001	ttgacagcta	tgatccagat	gtcagaatta	ttgagaacca	tggaactgag	gatgagttgc
24061	caaattattg	ctttcctctt	ggtggaattg	ggattactga	cacttttcaa	gctgttaaaa
24121	caactgctgc	taacggggac	caaggcaata	ctacctggca	aaaagattca	acatttgcag
24181	aacgcaatga	aataggggtg	ggaaataact	ttgccatgga	aattaacctg	aatgccaacc
24241	tatggagaaa	tttcctttac	tccaatattg	cgctgtacct	gccagacaag	ctaaaataca
24301	accccaccaa	tgtggaaata	tctgacaacc	ccaacaccta	cgactacatg	aacaagcgag
24361	tggtggctcc	tgggcttgta	gactgctaca	ttaaccttgg	ggcgcgctgg	tctctggact
24421	acatggacaa	cgttaatccc	tttaaccacc	accgcaatgc	gggcctgcgt	taccgctcca
24481	tgttgttggg	aaacggccgc	tacgtgccct	ttcacattca	ggtgccccaa	aagttttttg
24541	ccattaaaaa	cctcctcctc	ctgccaggct	catacacata	tgaatggaac	ttcaggaagg
24601	atgttaacat	ggttctgcag	agctctctgg	gaaacgacct	tagagttgac	ggggctagca
24661	ttaagtttga	cagcatttgt	ctttacgcca	ccttcttccc	catggcccac	aacacggcct
24721	ccacgctgga	agccatgctc	agaaatgaca	ccaacgacca	gtcctttaat	gactaccttt
24781	ccgccgccaa	catgctatat	cccatacccg	ccaacgccac	caacgtgccc	atctccatcc
24841	catcgcgcaa	ctgggcagca	tttcgcggtt	gggccttcac	acgcttgaag	acaaaggaaa
24901	ccccttccct	gggatcaggc	tacgaccctt	actacaccta	ctctggctcc	ataccatacc
24961	ttgacggaac	cttctatctt	aatcacacct	ttaagaaggt	ggccattact	tttgactctt
25021	ctgttagctg	gccgggcaac	gaccgcctgc	ttactcccaa	tgagtttgag	attaagcgct
25081	cagttgacgg	ggagggctat	aacgtagctc	agtgcaacat	gacaaaggac	tggttcctag
25141	tgcagatgtt	ggccaactac	aatattggct	accagggctt	ctacattcca	gaaagctaca
25201	aagaccgcat	gtactcgttc	ttcagaaact	tccagcccat	gagccggcaa	gtggtggacg
25261	atactaaata	caaagattat	cagcaggttg	gaattatcca	ccagcataac	aactcaggct
25321	tcgtaggcta	cctcgctccc	accatgcgcg	agggacaagc	ttaccccgct	aatgttccct
25381	acccactaat	aggcaaaacc	gcggttgata	gtattaccca	gaaaaagttt	ctttgcgacc
25441	gcaccctgtg	gcgcatcccc	ttctccagta	actttatgtc	catgggtgcg	ctcacagacc
25501	tgggccaaaa	ccttctctac	gcaaactccg	cccacgcgct	agacatgacc	tttgaggtgg
25561	atcccatgga	cgagcccacc	cttctttatg	ttttgtttga	agtctttgac	gtggtccgtg
25621	tgcaccagcc	gcaccgcggc	gtcatcgaga	ccgtgtacct	gcgcacgccc	ttctcggccg
25681	gcaacgccac	aacataaaga	agcaagcaac	atcaacaaca	gctgccgcca	tgggctccag
25741	tgagcaggaa	ctgaaagcca	ttgtcaaaga	tcttggttgt	gggccatatt	ttttgggcac
25801	ctatgacaag	cgcttcccag	gctttgtttc	cccacacaag	ctcgcctgcg	ccatagttaa
25861	cacggccggt	cgcgagactg	ggggcgtaca	ctggatggcc	tttgcctgga	accegegete
25921	aaaaacatgc	tacctctttg	agccctttgg	cttttctgac	caacgtctca	agcaggttta
25981	ccagtttgag	tacgagtcac	tcctgcgccg	tagcgccatt	gcctcttccc	ccgaccgctg
26041	tataacgctg	gaaaagtcca	cccaaagcgt	gcaggggccc	aactcggccg	cctgtggcct
26101	attctgctgc	atgtttctcc	acgcctttgc	caactggccc	caaactccca	tggatcacaa
26161	cccaccatg	aaccttatta	ccggggtacc	caactccatg	cttaacagtc	cccaggtaca
26221	gcccaccctg	cgccgcaacc	aggaacagct	ctacagcttc	ctggagcgcc	actcgcccta
26281	cttccgcagc	cacagtgcgc	aaattaggag	cgccacttct	ttttgtcact	tgaaaaacat
26341	gtaaaaataa	tgtactagga	gacactttca	ataaaggcaa	atgtttttat	ttgtacactc
26401	tcgggtgatt	atttaccccc	accettgeeg	tctgcgccgt	ttaaaaatca	aaggggttct

FIG. 41

						++ actactac
26461	gccgcgcatc	gctatgcgcc	actggcaggg	acacgttgcg	atactggtgt	cagagatas
26521	acttaaactc	aggcacaacc	atccgcggca	gctcggtgaa	gttttcactc	cacaggerge
26581	gcaccatcac	caacgcgttt	agcaggtcgg	gcgccgatat	cttgaagtcg	cagingggge
26641	ctccgccctg	cgcgcgcgag	ttgcgataca	cagggttaca	gcactggaac	actateageg
26701	ccgggtggtg	cacgctggcc	agcacgctct	tgtcggagat	cagatccgcg	tecaggieei
26761	ccacattact	cagggcgaac	ggagtcaact	ttggtagctg	ccttcccaaa	aagggtgcat
26821	gcccaggett	tgagttgcac	tcgcaccgta	gtggcatcag	aaggtgaccg	tgcccagtct
26881	gggcgttagg	atacagcgcc	tgcatgaaag	ccttgatctg	cttaaaagcc	acctgageet
26941	ttacacette	agagaagaac	atgccgcaag	acttgccgga	aaactgattg	gccggacagg
27001	ccgcgtcatg	cacacaqcac	cttgcgtcgg	tgttggagat	ctgcaccaca	ttteggeece
27061	accognition	cacgatettg	gccttgctag	actgctcctt	cagcgcgcgc	tgeeegtttt
27121	cgctcgtcac	atccatttca	atcacgtgct	ccttatttat	cataatgctc	cegtgtagae
27181	acttaagctc	gccttcgatc	tcagcgcagc	ggtgcagcca	caacgcgcag	cccgtgggct
27241	cataatactt	gtaggttacc	tctgcaaacg	actgcaggta	cgcctgcagg	aategeeeca
27301	tratrotrac	aaaggtcttg	ttgctggtga	aggtcagctg	caacccgcgg	tgeteetegt
27361	ttagccaggt	cttgcatacg	gccgccagag	cttccacttg	gtcaggcagt	agettgaagt
27421	ttgcctttag	atcottatcc	acgtggtact	tgtccatcaa	cgcgcgcgca	geeteeatge
27481	ccttctccca	cacagacaca	atcggcaggc	tcagcgggtt	tatcaccgtg	CTTTCACTTL
27541	ccacttcact	ggactcttcc	ttttcctctt	gcatccgcat	accccgcgcc	actgggtcgt
27501	cttcattcad	ccaccacacc	gtgcgcttac	ctcccttgcc	gtgcttgatt	agcaccggtg
27661	ggttgctgaa	acccaccatt	tgtagcgcca	catcttctct	ttcttcctcg	ctgtccacga
27721	tcacctctgg	agatagcaga	cgctcgggct	tgggagaggg	gcgcttcttt	ttetttigg
27781	acgcaatggc	caaatccqcc	gtcgaggtcg	atggccgcgg	gctgggtgtg	egeggeacea
27841	gcgcatcttg	tgacgagtct	tcttcgtcct	cggactcgag	acgccgcctc	ageegetttt
27901	ttaaaaacac	acaaaaaagc	ggcggcgacg	gcgacgggga	cgagacgtcc	cecargging
27961	atagacatea	cgccgcaccg	cgtccgcgct	cgggggtggt	ttcgcgctgc	feetetteee
28021	gactggccat	ttccttctcc	tataggcaga	aaaagatcat	ggagtcagtc	gagaaggagg
28081	acageetaae	cacccccttt	gagttcgcca	ccaccgcctc	caccgatgcc	gccaacgcgc
28141	ctaccacctt	ccccatcaaa	gcacccccgc	ttgaggagga	ggaagtgatt	atcgagcagg
28201	acccaggitt	totaaocgaa	qacgacgaag	atcgctcagt	accaacagag	gataaaaagc
28261	aacaccacca	cgacgcagag	gcaaacgagg	aacaagtcgg	gcggggggac	caaaggcatg
28321	gcgactacct	agatgtggga	gacgacgtgc	tgttgaagca	tctgcagcgc	cagtgcgcca
28381	ttatctgcga	cacattacaa	gagcgcagcg	atgtgcccct	cgccatagcg	gatgtcagcc
20301	ttacctacaa	acccaccto	ttctcaccgc	gcgtaccccc	caaacgccaa	gaaaacggca
28501	catocoacc	caacccgcgc	ctcaacttct	accccgtatt	tgccgtgcca	gaggtgcttg
28561	ccacctatca	catctttttc	caaaactgca	agatacccct	atcctgccgt	gccaaccgca
20501	accasacas	caaggaggtg	accttacaac	agggcgctgt	catacctgat	atcgcctcgc
20021	tcaacaaat	accasasate	tttgagggtc	ttggacgcga	cgagaagcgc	gcggcaaacg
20001	ctctccaaca	agaaaacagc	gaaaatgaaa	gtcactgtgg	agtgctggtg	gaacttgagg
20741	atascascac	acacctaacc	gtgctgaaac	gcagcatcga	ggtcacccac	tttgcctacc
20001	caccacttaa	cctaccccc	aaggttatga	gcacagtcat	gagcgagctg	atcgtgcgcc
20001	atacacasc	cctagagaga	gatgcaaact	tgcaagaaca	aaccgaggag	ggcctacccg
20321	grycacyacc	traccarctr	acacactaac	ttgagacgcg	cgagcctgcc	gacttggagg
20041	cagicggcga	actaataata	accacaatac	ttgttaccgt	ggagcttgag	tgcatgcagc
29041	agegacgeaa	traccorran	atacaacaca	agctagagga	aacgttgcac	tacacctttc
29101	ggttetttgt	catacaccaa	acctacaaaa	tttccaacgt	ggagctctgc	aacctggtct
29101	gccagggcta	cgtgcgccag	geeegeaaaa	traggraaaa	cgtgcttcat	tccacgctca
29221	cctaccttgg	aattttgcat	tacatccaca	actocottta	cttatttctg	toctacacct
29281	agggcgaggc	gegeegegae	tacgecegeg	acctagaaa	gcgcaacctg	aaggagctgc
29341	ggcaaacggc	catgggcgtg	trassass	tatoracoro	gcgcaacctg	cactccataa
29401	agaagctgct	aaagcaaaac	atattaca:	aacgcctcct	cttcaacgag	caacagggtc
29461	ccgcgcacct	ggcggacact	accelected	aaaactttac	taaaaccctg	ctagagcgtt
29521	tgccagactt	caccagtcaa	agcatgttgc	ttactage	gaactttatc	attaagtacc
29581	caggaattct	gcccgccacc	taaataat	actacattat	ctttgtgccc	aactacctto
29641	gtgaatgccc	teegeegett	Lggggtcact	gotaccetce	gcagctagcc	tatcactate
29701	cctaccactc	cgacatcatg	gaagacgtga	geggegaegg	cctactggag	cyccactyce

FIG. 4J

29761	gctgcaacct	atgcaccccg	caccgctccc	tggtctgcaa	ttcacaactg	cttagcgaaa
29821	gtcaaattat	cggtaccttt	gagctgcagg	gtccctcgcc	tgacgaaaag	tccgcggctc
29881	cggggttgaa	actcactccg	gggctgtgga	cgtcggctta	ccttcgcaaa	tttgtacctg
29941	aggactacca	cgcccacgag	attaggttct	acgaagacca	atcccgcccg	ccaaatgcgg
30001	agettacege	ctgcgtcatt	acccagggcc	acatccttgg	ccaattgcaa	gccattaaca
30061	aagcccgcca	agagtttctg	ctacgaaagg	gacggggggt	ttacttggac	ccccagtccg
30121	gcgaggagct	caacccaatc	ccccgccgc	cgcagcccta	tcagcagccg	cgggcccttg
30181	cttcccagga	tggcacccaa	aaagaagctg	cagctgccgc	cgccgccacc	cacggacgag
30241	gaggaatact	gggacagtca	ggcagaggag	gttttggacg	aggaggagga	gatgatggaa
30301	gactgggaca	gcctagacga	ggaagcttcc	gaggccgaag	aggtgtcaga	cgaaacaccg
30361	tcaccctcgg	tcgcattccc	ctcgccggcg	ccccagaaat	cggcaaccgt	tcccagcatt
30421	gctacaacct	ccgctcctca	ggcgccgccg	gcactgcccg	ttcgccgacc	caaccgtaga
30481	tgggacacca	ctggaaccag	ggccggtaag	tctaagcagc	cgccgccgtt	agcccaagag
30541	caacaacagc	gccaaggcta	ccgctcgtgg	cgcgtgcaca	agaacgccat	agttgcttgc
30601	ttgcaagact	gtgggggcaa	catctccttc	gcccgccgct	ttcttctcta	ccatcacggc
		cccgtaacat				
30721	ggcggcagcg	gcagcaacag	cagcggccac	gcagaagcaa	aggcgaccgg	atagcaagac
30781	tctgacaaag	cccaagaaat	ccacagcggc	ggcagcagca	ggaggaggag	cactgcgtct
30841	ggcgcccaac	gaacccgtat	cgacccgcga	gcttagaaac	aggattttc	ccactctgta
30901	tgctatattt	caacagagca	ggggccaaga	acaagagctg	aaaataaaaa	acaggtctct
30961	gcgctccctc	acccgcagct	gcctgtatca	caaaagcgaa	gatcagcttc	ggcgcacgct
31021	ggaagacgcg	gaggctctct	tcagcaaata	ctgcgcgctg	actcttaagg	actagtttcg
31081	cgccctttct	caaatttaag	cgcgaaaact	acgtcatctc	cagcggccac	acccggcgcc
31141	agcacctgtc	gtcagcgcca	ttatgagcaa	ggaaattccc	acgccctaca	tgtggagtta
31201	ccagccacaa	atgggacttg	cggctggagc	tgcccaagac	tactcaaccc	gaataaacta
31261	catgagcgcg	ggaccccaca	tgatatcccg	ggtcaacgga	atccgcgccc	accgaaaccg
31321	aattctcctc	gaacaggcgg	ctattaccac	cacacctcgt	aataacctta	atccccgtag
31381	ttggcccgct	gccctggtgt	accaggaaag	tcccgctccc	accactgtgg	tacttcccag
31441	agacgcccag	gccgaagttc	agatgactaa	ctcaggggcg	cagcttgcgg	gcggctttcg
31501	tcacagggtg	cggtcgcccg	ggcagggtat	aactcacctg	aaaatcagag	ggcgaggtat
		gacgagtcgg				
		gctggccgct				
31681	ctcgtcctcg	gagccgcgct	ccggaggcat	tggaactcta	caatttattg	aggagttcgt
		tacttcaacc				
31801	tcccaacttt	gacgcggtaa	aagactcggc	ggacggctac	gactgaatga	ccagtggaga
31861	ggcagagcaa	ctgcgcctga	cacacctcga	ccactgccgc	cgccacaagt	gctttgcccg
31921	cggctccggt	gagttttgtt	actttgaatt	gcccgaagag	catatcgagg	gcccggcgca
31981	cggcgtccgg	ctcaccaccc	aggtagagct	tacacgtagc	ctgattcggg	agtttaccaa
32041	gcgccccctg	ctagtggagc	gggagcgggg	tccctgtgtt	ctgaccgtgg	tttgcaactg
32101	tcctaaccct	ggattacatc	aagatcttat	tccattcaac	taacaataaa	cacacaataa
32161	attacttact	taaaatcagt	cagcaaatct	ttgtccagct	tattcagcat	cacctccttt
32221	ccctcctccc	aactctggta	tttcagcagc	cttttagctg	cgaactttct	ccaaagtcta
32281	aatgggatgt	caaattcctc	atgttcttgt	ccctccgcac	ccactatctt	catattgttg
32341	cagatgaaac	gcgccagacc	gtctgaagac	accttcaacc	ctgtgtaccc	atatgacacg
		ctccaactgt				
		ccccggagt				
32521	ggcatgcttg	cgctaaaaat	gggcagcggc	ctgtccctgg	atcaggcagg	caaccttaca
		tcactgtttc				
		cccttacagt				
32701	gtggtctctg	acaacactct	taccatgcaa	tcacaagcac	cgctaaccgt	gcaagactca
32761	aaacttagca	ttgctaccaa	agagccactt	acagtgttag	atggaaaact	ggccctgcag
32821	acatcagccc	ccctctctgc	cactgataac	aacgccctca	ctatcactgc	ctcacctcct
32881	cttactactg	caaatggtag	tctggctgtt	accatggaaa	acccacttta	caacaacaat
32941	ggaaaacttg	ggctcaaaat	tggcggtcct	ttgcaagtgg	ccaccgactc	acatgcacta
33001	acactaggta	ctggtcaggg	ggttgcagtt	cataacaatt	tgctacatac	aaaagttaca

33061	ggcgcaatag	ggtttgatac	atctggcaac	atggaactta	aaactggaga	tggcctctat
33121	gragatageg	ccggtcctaa	ccaaaaacta	catattaatc	taaataccac	aaaaggcctt
33121	actiticaca	acaccgcaat	aacaattaac	gctggaaaag	ggttggaatt	tgaaacagac
332/11	tcctcaaaca	gaaatcccat	aaaaacaaaa	attogatcag	gcatacaata	taataccaat
33301	agagetatag	ttgcaaaact	tggaacaggc	ctcagttttg	acageteegg	agccataaca
33361	atagacaaca	taaacaatga	cagacttact	ctttqqacaa	caccagaccc	atccccaaat
33301	tacagaatta	cttcagataa	agactgcaag	ctaactctqq	cgctaacaaa	atgtggcagt
22421	casattttaa	gcactgtttc	agetttggca	gtatcaggta	atatogcctc	catcaatgga
225/1	actoreage	gtgtaaactt	ggttcttaga	tttgatgaca	acggagtgct	tatgtcaaat
33501	tcatcactoo	acaaacagta	ttggaacttt	agaaacgggg	actccactaa	cggtcaacca
22661	tacacttata	ctgttgggtt	tatoccaaac	ctaaaagctt	acccaaaaac	tcaaagtaaa
33701	actornana	gtaatattgt	tagccaggtg	tatcttaatq	gtgacaagtc	taaaccattg
22701	cattttacta	ttacgctaaa	tagaacagat	gaaaccaacc	aagtaagcaa	atactcaata
33/01	tasttasatt	ggtcctggaa	cagtagacaa	tacactaatq	acaaatttgc	caccaattcc
33001 3304T	tataccttct	cctacattgc	ccaggataa	agaatcgtga	acctgttgca	tgttatgttt
33301	cacaccctct	atttttcaat	tacagaaaat	ttcaagtcat	ttttcattca	gtagtatagc
33301	caacgtgttt	catagettat	actaatcacc	gtaccttaat	caaactcaca	gaaccctagt
34021	atteaacete	ccacctccct	cccaacacac	agagtacaca	atcetttete	cccggctggc
34001	atteaacety	atcatatcat	gggtaacaga	catattetta	ggtgttatat	tccacacggt
34141	ctcadacage	gccaaacgct	catcagtgat	gttaataaac	tececaaaca	gctcgcttaa
34201	ettestete	ctgtccagct	actagaccac	aggetgetgt	ccaacttgcg	gttgctcaac
34201	greeargrey	ggagaagtcc	acceptacat	aggetjeeje	tcataatcgt	gcatcaggat
34321	gggcggcgaa	tgctgcagca	acgcccacat	aaactoctoc	caccaccact	ccatcctaca
34381	agggeggtgg	atggcagtgg	tetesteace	catcattccc	accoccoca	gcataaggcg
34441	ggaatacaac	cgggcacagc	accececage	gatgateegt	aagtcagcac	agtaactgca
34501	cettgteete	acaatattgt	ttaaaatccc	acanthrosan	acactatate	caaagctcat
34561	gcacagtacc	acagaaccca	catagecete	ataccacaa	cocaontaga	ttaagtggcg
34621	ggcggggacc	acacacctgg	acataaacat	tacctcttt	agcatattat	aattcaccac
34681	acccctcata	catataaacc	totoattaaa	categococca	tccaccacca	tectaaacca
34/41	ctcccggtac	acctgcccgc	caactataca	ctacaaaaaa	ccaaaactaa	aacaatgaca
34801	gctggccaaa	caggactcgt	eaccatocat	catcatocto	gtcatgatat	caatgttggc
34861	geggagagee	cacacgtgca	tacacttcct	caccatgett	agetectece	gcgtcagaac
34921	acaacacagg	ggaacaaccc	attecteast	caggattaaat	ccacactec	agggaagacc
34981	catateceag	ctcacgttgt	accectgaac	agtattacat	tcaaacaaca	gcggatgatc
35041	tegeaegtaa	gtagcgcggg	tttctctc	agegeeacat	agacgatccc	tactgtacgg
35101	ctccagtatg	gragegeggg	atastattaa	testastate	ataccaaata	gaacgccgga
35161	agtgcgccga	tttcctgaag	accycyctyy	tagaagagata	acasacadat	ctacatetee
35221	cgtagtcata	cttagatcgc	totatataat	agregate	tatccactct	ctcaaagcat
35281	ggtctcgccg	cctggcttcg	aattatatat	agetycette	atacaccact	gccctgataa
35341	ccaggcgccc	cgcagaataa	ggttttatgt	accentec	acattcottc	tacaaatcac
35401	catccaccac	agcgggaaga	getacaccca	ccatatttt	ttttttattc	caaaagatta
35461	acacgggagg	caaaatgaag	atctattaac	taaacacact	ccctccat	gacataatca
35521	tccaaaacct	cadaatyaay	gataatggg	tttataaaat	ottocacaat	ggcttccaaa
35581	aactctacag	ccaaagaaca	gataatyyta	taaagac	accettcace	graatctcc
35641	aggcaaacgg	ccctcacgtc	taaguggaug	caaayyctaa	tetestetea	ccaccttctc
35701	tctataaaca	ttccagcacc	cccaaccatg	actocacca	tratagagat	ctactccaga
35761	aatatatctc	taagcaaatc	ccgaatatta	agteeggeea	cogcaaaaac	ctgctccaga
35821	gcgccctcca	cetteageet	caagcagcga	tonangatey	caaaaactca	ggttcctcac
35881	agacctgtat	aagattcaaa	ageggaacat	taacaaaat	accycyatec	cgtaggtccc
35941	ttcgcagggc	cagctgaaca	taatcgtgca	ggtetgeacg	gaccagegeg	gccacttccc
36001	cgccaggaac	catgacaaaa	gaacccacac	cgattatgac	acguatacto	ggagctatgc
36061	taaccagcgt	agccccgatg	Laagettgtt	gcatgggcgg	cyatataada	tgcaaggtgc
36121	tgctcaaaaa	atcaggcaaa	gcctcgcgca	aaaaagaaag	acategray	tcatgctcat
36181	gcagataaag	gcaggtaagc	tccggaacca	ccacagaaaa	agacaccatt	tttctctcaa
36241	acatgtctgc	gggtttctgc	ataaacacaa	aataaaataa	caaaaaaaca	tttaaacatt
36301	agaagcctgt	cttacaacag	gaaaaacaac	ccttataagc	acaayacyga	ctacggccat

36361	gccggcgtga	ccgtaaaaaa	actggtcacc	gtgattaaaa	agcaccaccg	acagctcctc
36421	ggtcatgtcc	ggagtcataa	tgtaagactc	ggtaaacaca	tcaggttgat	tcacatcggt
36481	cagtgctaaa	aagcgaccga	aatagcccgg	gggaatacat	acccgcaggc	gtagagacaa
36541	cattacagcc	cccataggag	gtataacaaa	attaatagga	gagaaaaaca	cataaacacc
36601	tgaaaaaccc	tcctgcctag	gcaaaatagc	accctcccgc	tccagaacaa	catacagcgc
36661	ttccacagcg	gcagccataa	cagtcagcct	taccagtaaa	aaagaaaacc	tattaaaaaa
36721	acaccactcg	acacggcacc	agctcaatca	gtcacagtgt	aaaaaagggc	caagtgcaga
36781	gcgagtatat	ataggactaa	aaaatgacgt	aacggttaaa	gtccacaaaa	aacacccaga
36841	aaaccgcacg	cgaacctacg	cccagaaacg	aaagccaaaa	aacccacaac	ttcctcaaat
36901	cgtcacttcc	gttttcccac	gttacgtcac	ttcccatttt	aagaaaacta	caattcccaa
36961	cacatacaag	ttactccgcc	ctaaaaccta	cgtcacccgc	cccgttccca	cgccccgcgc
37021	cacgtcacaa	actccacccc	ctcattatca	tattggcttc	aatccaaaat	aaggtatatt
37081	attgatgatg					

10	30	50
		GGCCTACTTGGTTGCATCATCACT
		++ GlyLeuLeuGlyCysIleIleThr
MetAlaProlleTh	ratatyr sergingininiargo 10	20
70	90	110
		GAGGTTCAGGTGGTTTCCACCGCA
		GluValGlnValValSerThrAla
Serbearmory	30	40
130	150	170
		TGTTGGACCGTTTACCATGGTGCT
		CysTrpThrValTyrHisGlyAla
	50	60
190	210	230
		ACCCAGATGTACACTAATGTGGAC
		ThrGlnMetTyrThrAsnValAsp
	. 70	80
250	270	290 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
		CGTTCCTTGACACCATGCACCTGT
		ArgSerLeuThrProCysThrCys
	90	100
		350
310	330 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	GACGTCATTCCGGTGCGCCGGCGG
		+
		AspVallleProValArgArgArg
	110	120
	200	410
370	390	410 GTCTCCTACTTGAAGGGCTCTTCG
		+
		ValSerTyrLeuLysGlySerSer
	130	140

FIG. 5A

170	47	450	430
		GCTCTGCCCTTCGGGGCAC	
	AlaValGlyIlePheArgA	uLeuCysProSerGlyHis/	GlyGlyProLe
160		150	
	F.3	510	
30		510	490
		rgcgaaggcggtggacttt(+	
		lAlaLysAlaValAspPhe\	
180	4111014101400111000	170	IMIAIGGIYVA
90	59	570	550
TCATTTCAAGTG	CCCCGGCCGTACCGCAGT	CTTCACGGACAACTCATCC	CGGTCTCCGGT
-+			
SerPheGlnVal	ProProAlaValProGlnS	lPheThrAspAsnSerSerI	ArgSerProVa
200		190	
			610
		CGCTCCCACTGGCAGCGGCA	
		+	•
220	ysserinibysvarrion.	AlaProThrGlySerGlyL 210	Alahisbeuhi
220		210	
10	710	690	670
TTAGGGTTTGGG		CAAGGTGCTCGTCCTCAATC	
		+	
LeuGlyPheGly	roSerValAlaAlaThrLe	LysValLeuValLeuAsnF	AlaGlnGlyTy
240		230	
*	770	750	730
		PAAGGCACACGGTATTGACC	
valarginille 260	roasnilearginrGlyVa	LysAlaHisGlyIleAspP 250	AlaTyrMetSe
260		250	
3.0	830	810	790
		CCCGTCACATACTCTACCT	
		+	
AspGlyGlyCys	yrGlyLysPheLeuAlaAs	ProValThrTyrSerThrT	ThrThrGlvAl:
280		270	

850	870	890
		CCATTCAACTGACTCGACTACA
		-++
SerGlyGlyAlaTyrAspII	ellelleCysAspGluCy 290	sHisSerThrAspSerThrThr
	250	
910	930	950
		GGCTGGAGCGCGGCTTGTCGTG
		-++ rAlaGlyAlaArgLeuValVal
lleLeuGlylleGlylniva.	310	320
	 -	
970	990	1010
		ACACCCAAACATCGAGGAGGTG
LeuAlaThrAlaThrProPr	oGlyservalinivalri 330	OHisProAsnIleGluGluVal
	330	
1030	1050	1070
		AGCCATCCCCATTGAAGCCATC
AlaLeuSerAsnThrGlyGl	ulleprophelyrGlyLy 350	sAlaIleProIleGluAlaIle/ 360
	330	
1090	1110	1130
		AGAAGTGCGACGAGCTCGCCGCA
ArgGlyGlyArgHisLeuIl	ePheCysHisSerLysL)	sLysCysAspGluLeuAlaAla 380
	370	•
1150	1170	1190
		ACCGGGGGCTCGATGTGTCCGTC
LysLeuSerGlyLeuGlyIl	eAsnAlaValAlaTyrT) 390	yrArgGlyLeuAspValSerVal 400
	390	400
1210	1230	1250
	CGTTGTCGTGGCAACAG	ACGCTCTGATGACGGGCTATACG
		++
IleProThrIleGlyAspVa	lValValValAlaThrA	spAlaLeuMetThrGlyTyrThr 420

	1310	1290	1270
CGACTTCAGC	ACATGTGTCACCCAGACAGT(CTCAGTGATCGACTGTAAC <i>i</i>	GGCGACTTTGAC
+	++		
lAspPheSer	ThrCysValThrGlnThrVal	oSerVallleAspCysAsn7	GlyAspPheAsp
440		430	
		1350	1330
	ACCGTGCCTCAAGACGCAGTG		
	ThrValProGlnAspAlaVal		LeuAspProThr
460		450	
	1430	1410	1390
ያልርጥ ር ርርርል	GAGGCATCTACAGGTTTGTG		
	argGlyIleTyrArgPheVal		
480		470	0 g g 0 j.
	1490	1470	1450
CCGGGCTGT	TCCTGTGTGAGTGCTATGAC	GGCATGTTCGATTCCTCGG	GAACGGCCCTCG
+	+	+	
AlaGlyCys	/alLeuCysGluCysTyrAsp	GlyMetPheAspSerSerV	GluArgProSer
500		490	
		1530	
	CGGTTAGGTTGCGGGCCTAC		
	erValArgLeuArgAlaTyr		AlaTrpTyrGlul
520		510	
	1610	. 1590	1570
ACAGGCCTC	AGTTCTGGGAGAGTGTCTTC		
	luPheTrpGluSerValPhe		
540		530	- · · · · ·
	1670	1650	1630
TTCCCCTAC	.CCAAGCAGGCAGGAGACAAC	GCACACTTCTTGTCCCAGA	ACCCACATAGAT
+	+	+ + -	
PheProTyr	hrLysGlnAlaGlyAspAsn	AlaHisPheLeuSerGlnT	ThrHisIleAsp/
560		550	

FIG. 5D

1690	1710	1730
		AGGCCCCACCTCCATCATGGGAT
		+
LeuValAlaTyrGlnAlaTh:	rValCysAlaArgAlaG	lnAlaProProProSerTrpAsp
	570	580
1750	1770	1790
		TGCACGGGCCAACACCCTTGCTG
		+
GlnMetTrpLysCysLeull		euHisGlyProThrProLeuLeu 600
	590	800
1810	1830	1850
		CCCACCCATAACCAAATACATC
		++
		hrHisProIleThrLysTyrIle
Tyrkigheddrykiavaror	610	620
	010	
1870	1890	1910
ATGGCATGCATGTCGGCTGA	CCTGGAGGTCGTCACTA	GCACCTGGGTGCTGGTGGGCGGA
		++
		erThrTrpValLeuValGlyGly
<u>-</u>	630	640
1930	1950	1970
GTCCTTGCAGCTCTGGCCGC	GTATTGCCTGACAACAG	GCAGTGTGGTCATTGTGGGTAGG
+		++
ValLeuAlaAlaLeuAlaAl	aTyrCysLeuThrThrG	lySerValValIleValGlyArg
	650	660
1990	2010	2030
ATTATCTTGTCCGGGAGGCC	GGCTATTGTTCCCGACA	GGGAGTTTCTCTACCAGGAGTTC
		++
IleIleLeuSerGlyArgPr	oAlaIleValProAspA	rgGluPheLeuTyrGlnGluPhe
	670	680
		2000
2050	2070	2090
		TCGAGCAGGGAATGCAGCTCGCC
		leGluGlnGlyMetGlnLeuAla
AspGlumetGluGluCysAl	asernisLeuriolyri 690	10010011101yMetGIIIDeuxia 700

FIG. 5E

O	2150	2130	2110
	CTGCAAACAGCCACCAAACAA		
	LeuGlnThrAlaThrLysGln		
720		710	
0	2210	2190	2170
	GCCCTTGAGACATTCTGGGCG		
	AlaLeuGluThrPheTrpAla		
0	2270	2250	2230
	CAGGCTTATCCACTCTGCCT		
	laGlyLeuSerThrLeuPro		
760	24017 204202 2 2 2 2 4 2 2 4 2 2 2 2 2 2 2 2	750	11pasiiriie11e3e161
0	2330	2310	2290
	CTATCACCAGCCGGCTCACC		
	erIleThrSerProLeuThr		
780		770	
o	2390	2370	2350
	TGGCTGCCCAACTCGCCCCC		
	alAlaAlaGlnLeuAlaPro		
800		790	
)	2450	2430	2410
	GTGCGGCTGTTGGCAGCATA		
	lyAlaAlaValGlySerIle		
. 820		810	
)	2510	2490	2470
	GAGCAGGAGTGGCCGGCGCGC		
	lyAlaGlyValAlaGlyAla		·
840		830	

FIG. 5F

PCT/US02/32512

2530	2550	2570
TTCAAGGTCATGAGCGGCG	AGATGCCCTCCACCGAG(BACCTGGTCAATCTACTTCCTGCC
		++
PheLysValMetSerGlyG		AspLeuValAsnLeuLeuProAla
	850	860
2590	2610	2630
ATCCTCTCTCCTGGCGCCCC	TGGTCGTCGGGGTCGTG?	rgtgcagcaatactgcgtcgacac
		++
		CysAlaAlaIleLeuArgArgHis
IleLeuSerProGlyAlaL		2ysAlaAlallebeuAlgAlgAlg.120
	870	880
2650	2670	2690
GTGGGTCCGGGAGAGGGGG	CTGTGCAGTGGATGAAC	CGGCTGATAGCGTTCGCCTCGCGG
		++
		ArgLeuIleAlaPheAlaSerArg
ValGIyProGlyGluGlyA		900
	890	900
2710	2730	2750
GGTAATCATGTTTCCCCCA	CGCACTATGTGCCTGAG	AGCGACGCGCAGCGCGTGTTACT
		+
		SerAspAlaAlaAlaArgValThr
GIYASHHISVAISEFFFOT		920
	910	920
2770	2790	2810
CAGATCCTCTCCAGCCTTA	CCATCACTCAGCTGCTG	AAAAGGCTCCACCAGTGGATTAAT
		+
		LysArgLeuHisGlnTrpIleAsn
GIMITELEUSEISEIDEGI		940
	930	340
2830	2850	2870
GAAGACTGCTCCACACCGT	GTTCCGGCTCGTGGCTA	AGGGATGTTTGGGACTGGATATGC
	.+	+
		ArgAspValTrpAspTrpIleCys
Glunspeysselimilio	950	960
	930	300
2890	2910	2930
ACGGTGTTGACTGACTTCA	AGACCTGGCTCCAGTCC	AAGCTCCTGCCGCAGCTACCGGGA
	.+	++
		LysLeuLeuProGlnLeuProGly
varnearmrnopriier	970	980
	310	500

FIG. 5G

2950	2970	2990
		CTGGCGGGGAGACGGCATCATG
		+++
ValProPhePheSercys	ginarggiylyrLysgiyva 990	lTrpArgGlyAspGlyIleMet 1000
	330	1000
3010	3030 .	3050
CAAACCACCTGCCCATGT	GGAGCACAGATCACCGGACA	TGTCAAAAACGGTTCCATGAGG
		-++
GlnThrThrCysProCys		sValLysAsnGlySerMetArg
	1010	1020
3070	3090	3110
		AACATTCCCCATCAACGCATAC
		-++
IleValGlyProLysThr	CysSerAsnThrTrpHisGl	yThrPheProIleAsnAlaTyr
	1030	1040
		24.50
	0.200	3170
		TTCTAGGGCGCTGTGGCGGGTG
		rSerArgAlaLeuTrpArgVal
	1050	1060
	3210	
		TTTCCACTACGTGACGGGCATG
		-++ pPheHisTyrValThrGlyMet
AlaAlaGluGluTyrval	Giuvaiiniaigvaigiyas 1070	prhenisiyivaiimidiyhet 1080
	2070	2000
3250	3270	3290
ACCACTGACAACGTAAAG	TGCCCATGCCAGGTTCCGGC	TCCTGAATTCTTCACGGAGGTG
		-++
ThrThrAspAsnValLys		aProGluPhePheThrGluVal
	1090	1100
3310	3330	3350
		GCCTCTCCTACGGGAGGAGGTT
	-+	•
AspGlyValArgLeuHis	ArgTyrAlaProAlaCysAr	gProLeuLeuArgGluGluVal
	1110	1120

3370	3390	3410
		ACAGCTACCATGCGAGCCCGAA
		-++ rGlnLeuProCysGluProGlu
Intried invalory Box	1130	1140
3430	3450	3470
		CTCCCACATCACAGCAGAAACG
		oSerHisIleThrAlaGluThr
1101157441.114441254	1150	1160
3490	3510	3530
		GGCCAGCTCTTCAGCTAGCCAG
AlaLysArgArgLeuAla		euAlaSerSerSerAlaSerGln 1180
	1170	1100
3550	3570	3590
		CCATGTCTCTCCGGACGCTGAC
		++
LeuSerAlaProSerLeu	LysAlaThrCysThrThrHi	sHisValSerProAspAlaAsp
	1190	1200
	200	2650
	3630	3650
		GCGGGAACATCACCCGCGTGGAG
		.yGlyAsnIleThrArgValGlu
Deal 1 Con To 1 Con T	1210	1220
3670	3690	3710
		CGCTTCGAGCGGAGGAGGATGAG
SerGluAsnLysValVal		roLeuArgAlaGluGluAspGlu 1240
	1230	1240
3730	3750	3770
		CAAGAAGTTCCCCGCAGCGATG
		+
ArgGluValSerValPro	AlaGluIleLeuArgLysSe	erLysLysPheProAlaAlaMet
	1250	1260

FIG. 51

3790	3810	3830
CCCATCTGGGCGCGCCCGG	ATTACAACCCTCCACTGT	TAGAGTCCTGGAAGGACCCGGAC
	+	++
ProlleTrpAlaArgProAs	pTyrAsnProProLeuL	euGluSerTrpLysAspProAsp
	1270	1280
3850	3870	3890
		CTATCAAGGCCCCTCCAATACCA
·	•	+
TyrValProProValValHi		roIleLysAlaProProIlePro
	1290	1300
2010	2020	3950
	3930	3950 CCTCCGTGTCTTCTGCCTTAGCG
		++
		erSerValSerSerAlaLeuAla
PIOPIONIGNIGDY SALGIN	1310	1320
	1510	2300
3970	3990	4010
		CGCCGTCGACAGCGGCACGGCG
		+
GluLeuAlaThrLysThrPh	eGlySerSerGluSerSe	erAlaValAspSerGlyThrAla
-	1330	1340
4030	4050	4070
ACCGCCCTTCCTGACCAGGC	CTCCGACGACGGTGACA	AAGGATCCGACGTTGAGTCGTAC
+		++
ThrAlaLeuProAspGlnAl	aSerAspAspGlyAspLy	ysGlySerAspValGluSerTyr
	1350	1360
4090	4110	4130
		CCGATCTCAGTGACGGGTCTTGG
SerSerMetProProLeuGl		roAspLeuSerAspGlySerTrp
	1370	1380
4150	4170	41.00
4150	4170	4190
	TAGTGAGGATGTCGTCTC	GCTGCTCAATGTCCTACACATGG
		sCysSerMetSerTyrThrTrp
SET THE AUTREF GENERALINE	aserGrunspvarvarcy	acyssermecseriyrimriip

FIG. 5J

4210	4230	4250
ACAGGCGCCTTGATCACG	CATGCGCTGCGGAGGAAAG	CAAGCTGCCCATCAACGCGTTG
	-+	-++
ThrGlyAlaLeuIleThr	ProCysAlaAlaGluGluSe	rLysLeuProIleAsnAlaLeu
	1410	1420
4270	4290	4310
AGCAACTCTTTGCTGCGC	CACCATAACATGGTTTATGC	CACAACATCTCGCAGCGCAGGC
		-++
SerAsnSerLeuLeuArgi	HisHisAsnMetValTyrAl	aThrThrSerArgSerAlaGly
	1430	1440
		•
4330	4350	4370
		CCTGGACGACCACTACCGGGAC
LeuArgGlnLysLysVal		lLeuAspAspHisTyrArgAsp
	1450	1460
	4440	4430
4390	4410	
GTGCTCAAGGAGATGAAG	GCGAAGGCGTCCACAGTTAA	AGGCTAAACTCCTATCCGTAGAG
	-+	/sAlaLysLeuLeuSerValGlu
ValLeuLysGlumetLys	AlaLyskiaseiimivaiby 1470	1480
	1470	
4450	4470	4490
	• •	CCAAGTTTGGCTATGGGGCAAAG
		++
		erLysPheGlyTyrGlyAlaLys
Gimilacy oby bottom	1490	1500
4510	4530	4550
GACGTCCGGAACCTATCC	AGCAAGGCCGTTAACCACA	TCCACTCCGTGTGGAAGGACTTG
	-+	++
AspValArgAsnLeuSer	:SerLysAlaValAsnHisI	leHisSerValTrpLysAspLeu
_	1510	1520
4570	4590	4610
CTGGAAGACACTGTGACA	ACCAATTGACACCACCATCA	TGGCAAAAAATGAGGTTTTCTGT
		+
LeuGluAspThrValThr	ProlleAspThrThrlleM	etAlaLysAsnGluValPheCys
	1530	1540

FIG. 5K

4630	4650	4670
		TATCGTATTCCCAGATCTGGGA
		-++
ValGlnProGluLysGlyG		culleValPheProAspLeuGly
	1550	1560
4690	4710	4730
		CTCCACCCTTCCTCAGGTCGTG
		1SerThrLeuProGlnValVal
valArgvarcysGrubysh	1570	1580
	-3	
4750	4770	4790
ATGGGCTCCTCATACGGAT	TCCAGTACTCTCCTGGGCA	GCGAGTCGAGTTCCTGGTGAAT
	+	-++
MetGlySerSerTyrGlyP	heGlnTyrSerProGlyGl	nArgValGluPheLeuValAsn
	1590	1600
4810	4830	4850
		TGACACTCGCTGTTTCGACTCA
·		rAspThrArgCysPheAspSer
Thrippysserbysbysk	.1610	1620
	.1010	1020
4870	4890	4910
		TTACCAATGTTGTGACTTGGCC
	+	-++
ThrValThrGluAsnAspI	leArgValGluGluSerIl	eTyrGlnCysCysAspLeuAla
	1630	1640
4930	4950	4970
		GCTTTATATCGGGGGTCCTCTG
		-+
ProGluAlaArgGinAlal	1650	gLeuTyrIleGlyGlyProLeu 1660
	1030	1000
4990	5010	5030
		CCGCGCGAGCGGCGTGCTGACG
	+	
ThrAsnSerLysGlyGlnA:	snCysGlyTyrArgArgCy	sArgAlaSerGlyValLeuThr
	1670	1680

FIG. 5L

5050	5070	5090			
ACTAGCTGCGGTAACAC	CCTCACATGTTACTTGAAGGC	CTCTGCAGCCTGTCGAGCTGCG			
	+	-++			
ThrSerCysGlyAsnTh	rLeuThrCysTyrLeuLysAla	aSerAlaAlaCysArgAlaAla			
	1690	1700			
5110	5130	5150			
AAGCTCCAGGACTGCAC	GATGCTCGTGAACGGAGACGA(CCTTGTCGTTATCTGTGAAAGC			
	+				
LysLeuGlnAspCysTh	rMetLeuValAsnGlyAspAsp	pLeuValValIleCysGluSer			
	1710	1720			
5170	5190	5210			
	CGCGGCGAGCCTACGAGTCTTC				
	+				
AlaGlyThrGlnGluAs	pAlaAlaSerLeuArgValPho	eThrGluAlaMetThrArgTyr			
	1730	1740			
5230		5270			
		CTTGGAGCTGATAACATCATGT			
	+				
SerAlaProProGlyAs	pProProGlnProGluTyrAsp	pLeuGluLeuIleThrSerCys			
	1750	1760			
5290	5310	5330			
		AAGGGTGTACTACCTCACCCGT			
		-++			
SerSerAsnValSerVa		sArgValTyrTyrLeuThrArg			
	1770	1780			
5350	5370	5390			
		AGCTAGACACACTCCAGTTAAC			
AspProThrThrProLe		rAlaArgHisThrProValAsn 1800			
	1790	1800			
	F 4 2 2	EAEO			
5410	5430	5450			
	${\tt TCCTGGCTAGGCAACATTATCATGTATGCGCCCACTTTGTGGGCAAGGATGATTCTGATG}$				
SerTrpLeuGlyAsnIleIleMetTyrAlaProThrLeuTrpAlaArgMetIleLeuMet					
SerTrpLeuGlyAsnIl					
	1810	1820			

FIG. 5M

5470	5490	5510
		TTGAAAAAGCCCTGGACTGCCAG
		++ euGluLysAlaLeuAspCysGln
ThraispherneserileLe		
	1830	1840
5530	5550	5570
		PACCTCAGATCATTGAACGACTC
		+++ euProGlnIleIleGluArgLeu
TieryrGiyAracysTyrSe	1850	1860
	1830	1000
5590	5610	5630
		CAGGTGAGATCAATAGGGTGGCT
		++ coGlyGluIleAsnArgValAla
	1870	1880
	10.0	
5650	5670	5690
		rctggagacatcgggccaggagc
		++ alTrpArgHisArgAlaArgSer
	1890	1900
5710	5730	5750
		CACTTGTGGCAAGTACCTCTTC
		aThrCysGlyLysTyrLeuPhe
vainightanigheabeabe.	1910	1920
5770	5790	5810
		CCCGGCTGCGTCCCAGCTGGAC
·		.eProAlaAlaSerGlnLeuAsp
ASHILDAIAVAIDJOINIOJ	1930	1940
	2000	
5830	5850	5870
		CATATATCACAGCCTGTCTCGT
		plleTyrHisSerLeuSerArg
pensereralitips Hever Hrs	agiyiyiseigiygiyas 1950	prieryrnisserbedsernig

FIG. 5N

5890	5910	5930						
GCCCGACCCCGCTGGTT	GCCCGACCCCGCTGGTTCATGCTGTGCCTACTCCTACTTTCTGTAGGGGTAGGCATCTAC							
		++						
AlaArgProArgTrpPl	AlaArgProArgTrpPheMetLeuCysLeuLeuLeuLeuSerValGlyValGlyIleTyr							
	1970	1980						
•								
5950 5955								
CTGCTCCCCAACCGA	(SEQ. ID. NO. 5)							
LeuLeuProAsnArg	(SEQ. ID. NO. 6)							
1985								

1	TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG
51	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101	TCAGGGCGCG	TCAGCGGGTG	TTGGCGGGTG	TCGGGGCTGG	CTTAACTATG
151	CGGCATCAGA	GCAGATTGTA	CTGAGAGTGC	ACCATATGCG	GTGTGAAATA
201	CCGCACAGAT	GCGTAAGGAG	AAAATACCGC	ATCAGATTGG	CTATTGGCCA
251	TTGCATACGT	TGTATCCATA	TCATAATATG	TACATTTATA	TTGGCTCATG
301	TCCAACATTA	CCGCCATGTT	GACATTGATT	ATTGACTAGT	TATTAATAGT
351	AATCAATTAC	GGGGTCATTA	GTTCATAGCC	CATATATGGA	GTTCCGCGTT
401	ACATAACTTA	CGGTAAATGG	CCCGCCTGGC	TGACCGCCCA	ACGACCCCCG
451	CCCATTGACG	TCAATAATGA	CGTATGTTCC	CATAGTAACG	CCAATAGGGA
501	CTTTCCATTG	ACGTCAATGG	GTGGAGTATT	TACGGTAAAC	TGCCCACTTG
551	GCAGTACATC	AAGTGTATCA	TATGCCAAGT	ACGCCCCTA	TTGACGTCAA
601	TGACGGTAAA	TGGCCCGCCT	GGCATTATGC	CCAGTACATG	ACCTTATGGG
651	ACTTTCCTAC	TTGGCAGTAC	ATCTACGTAT	TAGTCATCGC	${\tt TATTACCATG}$
701	GTGATGCGGT	TTTGGCAGTA	CATCAATGGG	CGTGGATAGC	GGTTTGACTC
751	ACGGGGATTT	CCAAGTCTCC	ACCCCATTGA	CGTCAATGGG	AGTTTGTTTT
801	GGCACCAAAA	TCAACGGGAC	TTTCCAAAAT	GTCGTAACAA	CTCCGCCCCA
851	TTGACGCAAA	TGGGCGGTAG	GCGTGTACGG	${\tt TGGGAGGTCT}$	ATATAAGCAG
901	AGCTCGTTTA	GTGAACCGTC	AGATCGCCTG	GAGACGCCAT	CCACGCTGTT
951	TTGACCTCCA	TAGAAGACAC	CGGGACCGAT	CCAGCCTCCG	CGGCCGGGAA
1001	CGGTGCATTG	GAACGCGGAT	TCCCCGTGCC	AAGAGTGACG	TAAGTACCGC
1051	CTATAGACTC	TATAGGCACA	CCCCTTTGGC	TCTTATGCAT	GCTATACTGT
1101	TTTTGGCTTG	GGGCCTATAC	ACCCCCGCTT	CCTTATGCTA	TAGGTGATGG
1151	TATAGCTTAG	CCTATAGGTG	${\tt TGGGTTATTG}$	ACCATTATTG	ACCACTCCCC
1201	TATTGGTGAC	${\tt GATACTTTCC}$	ATTACTAATC	CATAACATGG	CTCTTTGCCA
1251	CAACTATCTC	${\tt TATTGGCTAT}$	ATGCCAATAC	TCTGTCCTTC	AGAGACTGAC
1301	ACGGACTCTG	${\tt TATTTTTACA}$	GGATGGGGTC	CCATTTATTA	TTTACAAATT
1351	CACATATACA	ACAACGCCGT	CCCCCGTGCC	CGCAGTTTTT	ATTAAACATA
1401	GCGTGGGATC	TCCACGCGAA	TCTCGGGTAC	GTGTTCCGGA	CATGGGCTCT
1451	TCTCCGGTAG	CGGCGGAGCT	TCCACATCCG	AGCCCTGGTC	CCATGCCTCC
1501	AGCGGCTCAT	GGTCGCTCGG	CAGCTCCTTG	CTCCTAACAG	TGGAGGCCAG
1551	ACTTAGGCAC	AGCACAATGC	CCACCACCAC	CAGTGTGCCG	CACAAGGCCG
1601	TGGCGGTAGG	GTATGTGTCT	GAAAATGAGC	GTGGAGATTG	GGCTCGCACG
1651	GCTGACGCAG	ATGGAAGACT	TAAGGCAGCG	GCAGAAGAAG	ATGCAGGCAG
1701	CTGAGTTGTT	GTATTCTGAT	AAGAGTCAGA	GGTAACTCCC	GTTGCGGTGC
1751	TGTTAACGGT	GGAGGGCAGT	GTAGTCTGAG	CAGTACTCGT	TGCTGCCGCG
1801	CGCGCCACCA	GACATAATAG	CTGACAGACT	AACAGACTGT	TCCTTTCCAT
1851	GGGTCTTTTC	TGCAGTCACC	GTCCTTAGAT	CTAGGTACCA	GATATCAGAA
1901	TTCAGTCGAC	AGCGGCCGCG	ATCTGCTGTG	CCTTCTAGTT	GCCAGCCATC
1951	TGTTGTTTGC	CCCTCCCCG	TGCCTTCCTT	GACCCTGGAA	GGTGCCACTC
2001	CCACTGTCCT	TTCCTAATAA	AATGAGGAAA	TTGCATCGCA	TTGTCTGAGT
2051	AGGTGTCATT	CTATTCTGGG	GGGTGGGGTG	GGGCAGGACA	GCAAGGGGGA

FIG. 6A

2101	GGATTGGGAA	GACAATAGCA	GGCATGCTGG	GGATGCGGTG	GGCTCTATGG
2151	CCGCTGCGGC	CAGGTGCTGA	AGAATTGACC	CGGTTCCTCC	TGGGCCAGAA
2201	AGAAGCAGGC	ACATCCCCTT	CTCTGTGACA	CACCCTGTCC	ACGCCCCTGG
2251	TTCTTAGTTC	CAGCCCCACT	CATAGGACAC	TCATAGCTCA	GGAGGGCTCC
2301	GCCTTCAATC	CCACCCGCTA	AAGTACTTGG	AGCGGTCTCT	CCCTCCCTCA
2351	TCAGCCCACC	AAACCAAACC	TAGCCTCCAA	GAGTGGGAAG	AAATTAAAGC
2401	AAGATAGGCT	ATTAAGTGCA	GAGGGAGAGA	AAATGCCTCC	AACATGTGAG
2451	GAAGTAATGA	GAGAAATCAT	AGAATTTCTT	CCGCTTCCTC	GCTCACTGAC
2501	TCGCTGCGCT	CGGTCGTTCG	GCTGCGGCGA	GCGGTATCAG	CTCACTCAAA
2551	GGCGGTAATA	CGGTTATCCA	CAGAATCAGG	GGATAACGCA	GGAAAGAACA
2601	TGTGAGCAAA	AGGCCAGCAA	AAGGCCAGGA	ACCGTAAAAA	GGCCGCGTTG
2651	CTGGCGTTTT	TCCATAGGCT	CCGCCCCCT	GACGAGCATC	ACAAAAATCG
2701	ACGCTCAAGT	CAGAGGTGGC	GAAACCCGAC	AGGACTATAA	AGATACCAGG
2751	CGTTTCCCCC	TGGAAGCTCC	CTCGTGCGCT	CTCCTGTTCC	GACCCTGCCG
2801	CTTACCGGAT	ACCTGTCCGC	CTTTCTCCCT	TCGGGAAGCG	TGGCGCTTTC
2851	TCATAGCTCA	CGCTGTAGGT	ATCTCAGTTC	GGTGTAGGTC	GTTCGCTCCA
2901	AGCTGGGCTG	TGTGCACGAA	CCCCCGTTC	AGCCCGACCG	CTGCGCCTTA
2951	TCCGGTAACT	ATCGTCTTGA	GTCCAACCCG	GTAAGACACG	ACTTATCGCC
3001	ACTGGCAGCA	GCCACTGGTA	ACAGGATTAG	CAGAGCGAGG	TATGTAGGCG
3051	GTGCTACAGA	GTTCTTGAAG	TGGTGGCCTA	ACTACGGCTA	CACTAGAAGA
3101	ACAGTATTTG	GTATCTGCGC	TCTGCTGAAG	CCAGTTACCT	TCGGAAAAAG
3151	AGTTGGTAGC	TCTTGATCCG	GCAAACAAAC	CACCGCTGGT	AGCGGTGGTT
3201	TTTTTGTTTG	CAAGCAGCAG	ATTACGCGCA	GAAAAAAAGG	ATCTCAAGAA
3251	GATCCTTTGA	TCTTTTCTAC	GGGGTCTGAC	GCTCAGTGGA	ACGAAAACTC
3301	ACGTTAAGGG	ATTTTGGTCA	TGAGATTATC	AAAAAGGATC	TTCACCTAGA
3351	TCCTTTTAAA	TTAAAAATGA	AGTTTTAAAT	CAATCTAAAG	TATATATGAG
3401	TAAACTTGGT	CTGACAGTTA	CCAATGCTTA	ATCAGTGAGG	CACCTATCTC
3451	AGCGATCTGT	CTATTTCGTT	CATCCATAGT	TGCCTGACTC	GGGGGGGGG
3501	GGCGCTGAGG	TCTGCCTCGT	GAAGAAGGTG	TTGCTGACTC	ATACCAGGCC
3551	TGAATCGCCC	CATCATCCAG	CCAGAAAGTG	AGGGAGCCAC	GGTTGATGAG
3601	AGCTTTGTTG	TAGGTGGACC	AGTTGGTGAT	TTTGAACTTT	TGCTTTGCCA
3651	CGGAACGGTC	TGCGTTGTCG	GGAAGATGCG	TGATCTGATC	CTTCAACTCA
3701	GCAAAAGTTC	GATTTATTCA	ACAAAGCCGC	CGTCCCGTCA	AGTCAGCGTA
3751	ATGCTCTGCC	AGTGTTACAA	CCAATTAACC	AATTCTGATT	AGAAAAACTC
3801	ATCGAGCATC	AAATGAAACT	GCAATTTATT	CATATCAGGA	TTATCAATAC
3851	CATATTTTTG	AAAAAGCCGT	TTCTGTAATG	AAGGAGAAAA	CTCACCGAGG
3901					TTCCGACTCG
3951					ATAAGGTTAT
4001					GAATGGCAAA
4051					CATTACGCTC
4101					CGTGATTGCG
4151	CCTGAGCGAG	ACGAAATACG	CGATCGCTGT	TAAAAGGAC	ATTACAAACA

FIG. 6B

4201	GGAATCGAAT	GCAACCGGCG	CAGGAACACT	GCCAGCGCAT	CAACAATATT
4251	TTCACCTGAA	TCAGGATATT	CTTCTAATAC	CTGGAATGCT	GTTTTCCCGG
4301	GGATCGCAGT	GGTGAGTAAC	CATGCATCAT	CAGGAGTACG	GATAAAATGC
4351	TTGATGGTCG	GAAGAGGCAT	AAATTCCGTC	AGCCAGTTTA	GTCTGACCAT
4401	CTCATCTGTA	ACATCATTGG	CAACGCTACC	TTTGCCATGT	TTCAGAAACA
4451	ACTCTGGCGC	ATCGGGCTTC	CCATACAATC	GATAGATTGT	CGCACCTGAT
4501	TGCCCGACAT	TATCGCGAGC	CCATTTATAC	CCATATAAAT	CAGCATCCAT
4551	GTTGGAATTT	AATCGCGGCC	${\tt TCGAGCAAGA}$	CGTTTCCCGT	TGAATATGGC
4601	TCATAACACC	${\tt CCTTGTATTA}$	CTGTTTATGT	AAGCAGACAG	TTTTATTGTT
4651	CATGATGATA	TATTTTTATC	TTGTGCAATG	TAACATCAGA	GATTTTGAGA
4701	CACAACGTGG	CTTTCCCCCC	CCCCCATTA	TTGAAGCATT	TATCAGGGTT
4751	ATTGTCTCAT	GAGCGGATAC	ATATTTGAAT	GTATTTAGAA	AAATAAACAA
.4801	ATAGGGGTTC	CGCGCACATT	TCCCCGAAAA	GTGCCACCTG	ACGTCTAAGA
4851	AACCATTATT	ATCATGACAT	TAACCTATAA	AAATAGGCGT	ATCACGAGGC
4901	CCTTTCGTC				

1	CATCATCAAT	AATATACCTT	ATTTTGGATT	GAAGCCAATA	TGATAATGAG	GGGGTGGAGT
61	TTGTGACGTG	GCGCGGGCG	TGGGAACGGG	GCGGGTGACG	TAGTAGTGTG	GCGGAAGTGT
121	GATGTTGTAA	GTGTGGCGGA	ACACATGTAA	GCGCCGGATG	TGGTAAAAGT	GACGTTTTTG
181	GTGTGCGCCG	GTGTACACGG	GAAGTGACAA	TTTTCGCGCG	GTTTTAGGCG	GATGTTGTAG
241	TAAATTTGGG	CGTAACCAAG	TAATATTTGG	CCATTTTCGC	GGGAAAACTG	AATAAGAGGA
301	AGTGAAATCT	GAATAATTCT	GTGTTACTCA	TAGCGCGTAA	TATTTGTCTA	GGGCCGCGGG
361	GACTTTGACC	GTTTACGTGG	AGACTCGCCC	AGGTGTTTTT	CTCAGGTGTT	TTCCGCGTTC
421	CGGGTCAAAG	TTGGCGTTTT	ATTATTATAG	TCAGCTGACG	CGCAGTGTAT	TTATACCCGG
481	TGAGTTCCTC	AAGAGGCCAC	TCTTGAGTGC	CAGCGAGTAG	AGTTTTCTCC	TCCGAGCCGC
541	TCCGACACCG	GGACTGAAAA	TGAGACATAT	TATCTGCCAC	GGAGGTGTTA	TTACCGAAGA
601	AATGGCCGCC	AGTCTTTTGG	ACCAGCTGAT	CGAAGAGGTA	CTGGCTGATA	ATCTTCCACC
661	TCCTAGCCAT	TTTGAACCAC	CTACCCTTCA	CGAACTGTAT	GATTTAGACG	TGACGGCCCC
721	CGAAGATCCC	AACGAGGAGG	CGGTTTCGCA	GATTTTTCCC	GAGTCTGTAA	TGTTGGCGGT
781	GCAGGAAGGG	ATTGACTTAT	TCACTTTTCC	GCCGGCGCCC	GGTTCTCCGG	AGCCGCCTCA
841	CCTTTCCCGG	CAGCCCGAGC	AGCCGGAGCA	GAGAGCCTTG	GGTCCGGTTT	CTATGCCAAA
901	CCTTGTGCCG	GAGGTGATCG	ATCTTACCTG	CCACGAGGCT	GGCTTTCCAC	CCAGTGACGA
961	CGAGGATGAA	GAGGGTGAGG	AGTTTGTGTT	AGATTATGTG	GAGCACCCCG	GGCACGGTTG
1021	CAGGTCTTGT	CATTATCACC	GGAGGAATAC	GGGGGACCCA	GATATTATGT	GTTCGCTTTG
1081	CTATATGAGG	ACCTGTGGCA	TGTTTGTCTA	CAGTAAGTGA	AAAATTATGG	GCAGTGGGTG
1141	ATAGAGTGGT	GGGTTTGGTG	TGGTAATTTT	TTTTTTAATT	TTTACAGTTT	TGTGGTTTAA
1201	AGAATTTTGT	ATTGTGATTT	TTTAAAAGGT	CCTGTGTCTG	AACCTGAGCC	TGAGCCCGAG
1261	CCAGAACCGG	AGCCTGCAAG	ACCTACCCGG	CGTCCTAAAT	TGGTGCCTGC	TATCCTGAGA
1321	CGCCCGACAT	CACCTGTGTC	TAGAGAATGC	AATAGTAGTA	CGGATAGCTG	TGACTCCGGT
1381	CCTTCTAACA	CACCTCCTGA	GATACACCCG	GTGGTCCCGC	TGTGCCCCAT	TAAACCAGTT
1441	GCCGTGAGAG	TTGGTGGGCG	TCGCCAGGCT	GTGGAATGTA	TCGAGGACTT	GCTTAACGAG
1501	TCTGGGCAAC	CTTTGGACTT	GAGCTGTAAA	CGCCCCAGGC	CATAAGGTGT	AAACCTGTGA
1561	TTGCGTGTGT	GGTTAACGCC	TTTGTTTGCT	GAATGAGTTG	ATGTAAGTTT	AATAAAGGGT
1621	GAGATAATGT	TTAACTTGCA	TGGCGTGTTA	AATGGGGCGG	GGCTTAAAGG	GTATATAATG
1681	CGCCGTGGGC	TAATCTTGGT	TACATCTGAC	CTCATGGAGG	CTTGGGAGTG	TTTGGAAGAT
1741	TTTTCTGCTG	TGCGTAACTT	GCTGGAACAG	AGCTCTAACA	GTACCTCTTG	GTTTTGGAGG
1801	TTTCTGTGGG	GCTCCTCCCA	GGCAAAGTTA	GTCTGCAGAA	TTAAGGAGGA	TTACAAGTGG
		AGCTTTTGAA				
		TCCAAGAGAA				
		TTGCTTTTTT				
2041	AGCGGGGGGT	ACCTGCTGGA	TTTTCTGGCC	ATGCATCTGT	GGAGAGCGGT	GGTGAGACAC
2101	AAGAATCGCC	TGCTACTGTT	GTCTTCCGTC	CGCCCGGCAA	TAATACCGAC	GGAGGAGCAA
		AAGCCAGGCG				
2221	GGCCTGGACC	CTCGGGAATG	AATGTTGTAC	AGGTGGCTGA	ACTGTTTCCA	GAACTGAGAC
		CATTAACGAG				
		TACAGAGGAG				
2401	CTGAGTGTGT	TACTTTTCAG	CAGATTAAGG	ATAATTGCGC	TAATGAGCTT	GATCTGCTGG
						GATGATTTTG

FIG. 7A

2521	AGGAGGCTAT	TAGGGTATAT	GCAAAGGTGG	CACTTAGGCC	AGATTGCAAG	TACAAGATTA
2581	GCAAACTTGT	AAATATCAGG	AATTGTTGCT	ACATTTCTGG	GAACGGGGCC	GAGGTGGAGA
2641	TAGATACGGA	GGATAGGGTG	GCCTTTAGAT	GTAGCATGAT	AAATATGTGG	CCGGGGGTGC
2701	TTGGCATGGA	CGGGGTGGTT	ATTATGAATG	TGAGGTTTAC	TGGTCCCAAT	TTTAGCGGTA
2761	CGGTTTTCCT	GGCCAATACC	AATCTTATCC	TACACGGTGT	AAGCTTCTAT	GGGTTTAACA
2821	ATACCTGTGT	GGAAGCCTGG	ACCGATGTAA	GGGTTCGGGG	CTGTGCCTTT	TACTGCTGCT
2881	GGAAGGGGGT	GGTGTGTCGC	CCCAAAAGCA	GGGCTTCAAT	TAAGAAATGC	CTGTTTGAAA
2941	GGTGTACCTT	GGGTATCCTG	TCTGAGGGTA	ACTCCAGGGT	GCGCCACAAT	GTGGCCTCCG
3001	ACTGTGGTTG	CTTTATGCTA	GTGAAAAGCG	TGGCTGTGAT	TAAGCATAAC	ATGGTGTGTG
3061	GCAACTGCGA	GGACAGGGCC	TCTCAGATGC	TGACCTGCTC	GGACGGCAAC	TGTCACTTGC
3121	TGAAGACCAT	TCACGTAGCC	AGCCACTCTC	GCAAGGCCTG	GCCAGTGTTT	GAGCACAACA
3181	TACTGACCCG	CTGTTCCTTG	CATTTGGGTA	ACAGGAGGG	GGTGTTCCTA	CCTTACCAAT
3241	GCAATTTGAG	TCACACTAAG	ATATTGCTTG	AGCCCGAGAG	CATGTCCAAG	GTGAACCTGA
3301	ACGGGGTGTT	TGACATGACC	ATGAAGATCT	GGAAGGTGCT	GAGGTACGAT	GAGACCCGCA
3361	CCAGGTGCAG	ACCCTGCGAG	TGTGGCGGTA	AACATATTAG	GAACCAGCCT	GTGATGCTGG
3421	ATGTGACCGA	GGAGCTGAGG	CCCGATCACT	TGGTGCTGGC	CTGCACCCGC	GCTGAGTTTG
3481	GCTCTAGCGA	TGAAGATACA	GATTGAGGTA	CTGAAATGTG	${\tt TGGGCGTGGC}$	TTAAGGGTGG
3541	GAAAGAATAT	ATAAGGTGGG	GGTCTCATGT	AGTTTTGTAT	${\tt CTGTTTTGCA}$	GCAGCCGCCG
3601	CCATGAGCGC	CAACTCGTTT	GATGGAAGCA	TTGTGAGCTC	ATATTTGÁCA	ACGCGCATGC
3661	CCCCATGGGC	CGGGGTGCGT	CAGAATGTGA	TGGGCTCCAG	CATTGATGGT	CGCCCCGTCC
3721	TGCCCGCAAA	CTCTACTACC	TTGACCTACG	AGACCGTGTC	TGGAACGCCG	TTGGAGACTG
3781	CAGCCTCCGC	CGCCGCTTCA	GCCGCTGCAG	CCACCGCCCG	${\tt CGGGATTGTG}$	ACTGACTTTG
3841	CTTTCCTGAG	CCCGCTTGCA	AGCAGTGCAG	CTTCCCGTTC	ATCCGCCCGC	GATGACAAGT
3901	TGACGGCTCT	TTTGGCACAA	TTGGATTCTT	${\tt TGACCCGGGA}$	ACTTAATGTC	GTTTCTCAGC
3961	AGCTGTTGGA	TCTGCGCCAG	CAGGTTTCTG	CCCTGAAGGC	$\mathtt{TTCCTCCCCT}$	CCCAATGCGG
4021	TTTAAAACAT	AAATAAAAAC	CAGACTCTGT	${\tt TTGGATTTGG}$	ATCAAGCAAG	TGTCTTGCTG
4081	TCTTTATTTA	GGGGTTTTGC	GCGCGCGGTA	GGCCCGGGAC	CAGCGGTCTC	GGTCGTTGAG
4141	GGTCCTGTGT	ATTTTTTCCA	GGACGTGGTA	AAGGTGACTC	TGGATGTTCA	GATACATGGG
4201	CATAAGCCCG	TCTCTGGGGT	GGAGGTAGCA	CCACTGCAGA	GCTTCATGCT	GCGGGGTGGT
4261	GTTGTAGATG	ATCCAGTCGT	AGCAGGAGCG	CTGGGCGTGG	TGCCTAAAAA	TGTCTTTCAG
4321	TAGCAAGCTG	ATTGCCAGGG	GCAGGCCCTT	GGTGTAAGTG	TTTACAAAGC	GGTTAAGCTG
4381	GGATGGGTGC	ATACGTGGGG	ATATGAGATG	CATCTTGGAC	TGTATTTTTA	GGTTGGCTAT
4441	GTTCCCAGCC	ATATCCCTCC	GGGGATTCAT	GTTGTGCAGA	ACCACCAGCA	CAGTGTATCC
4501	GGTGCACTTG	GGAAATTTGT	CATGTAGCTT	AGAAGGAAAT	GCGTGGAAGA	ACTTGGAGAC
4561	GCCCTTGTGA	CCTCCAAGAT	TTTCCATGCA	TTCGTCCATA	ATGATGGCAA	TGGGCCCACG
4621	GGCGGCGGCC	TGGGCGAAGA	TATTTCTGGG	ATCACTAACG	TCATAGTTGT	GTTCCAGGAT
4681	GAGATCGTCA	TAGGCCATTT	TTACAAAGCG	CGGGCGGAGG	GTGCCAGACT	GCGGTATAAT
4741	GGTTCCATCC	GGCCCAGGGG	CGTAGTTACC	CTCACAGATT	TGCATTTCCC	ACGCTTTGAG
4801	TTCAGATGGG	GGGATCATGT	CTACCTGCGG	GGCGATGAAG	AAAACCGTTT	CCGGGGTAGG
4861	GGAGATCAGC	TGGGAAGAAA	GCAGGTTCCT	AAGCAGCTGC	GACTTACCGC	AGCCGGTGGG
4921	CCCGTAAATC	ACACCTATTA	CCGGCTGCAA	CTGGTAGTTA	AGAGAGCTGC	AGCTGCCGTC
4981	ATCCCTGAGC	AGGGGGCCA	CTTCGTTAAG	CATGTCCCTG	ACTTGCATGT	TTTCCCTGAC

FIG. 7B

5041	CAAATCCGCC	AGAAGGCGCT	CGCCGCCCAG	CGATAGCAGT	TCTTGCAAGG	AAGCAAAGTT
					AGCGTTTGAC	
					CGATCCAGCA	
					GGTGCTCGTC	
					TAGTCTGGGT	
					GGCTGGTCCT	
					ATTTGACCAT	
					CCTTGGAGGA	
					CGAGAAATAC	
					ATTCCACGAG	
					TTTTGATGCG	
					GGCTGTCCGT	
5761	ACAGACTTGA	GAGGCCTGTC	CTCGAGCGGT	GTTCCGCGGT	CCTCCTCGTA	TAGAAACTCG
					AGGAGGCTAA	
					TGTGAAGACA	
					CCACGTGACC	
6001	GAAGGGGGC	TATAAAAGGG	GGTGGGGGCG	CGTTCGTCCT	CACTCTCTTC	CGCATCGCTG
6061	TCTGCGAGGG	CCAGCTGTTG	GGGTGAGTAC	TCCCTCTCAA	AAGCGGGCAT	GACTTCTGCG
6121	CTAAGATTGT	CAGTTTCCAA	AAACGAGGAG	GATTTGATAT	TCACCTGGCC	CGCGGTGATG
					TCTTTTTGTT	
					CGATGGAGCG	
6301	TTTTTGTCGC	GATCGGCGCG	CTCCTTGGCC	GCGATGTTTA	GCTGCACGTA	TTCGCGCGCA
6361	ACGCACCGCC	ATTCGGGAAA	GACGGTGGTG	CGCTCGTCGG	GCACTAGGTG	CACGCGCCAA
					CCTCTCCGCG	
					GCGGTAGTGG	
6541	GTCTCGTCCG	GGGGGTCTGC	GTCCACGGTA	AAGACCCCGG	GCAGCAGGCG	CGCGTCGAAG
6601	TAGTCTATCT	TGCATCCTTG	CAAGTCTAGC	GCCTGCTGCC	ATGCGCGGCC	GGCAAGCGCG
6661	CGCTCGTATG	GGTTGAGTGG	GGGACCCCAT	GGCATGGGGT	GGGTGAGCGC	GGAGGCGTAC
6721	ATGCCGCAAA	TGTCGTAAAC	GTAGAGGGGC	TCTCTGAGTA	TTCCAAGATA	TGTAGGGTAG
6781	CATCTTCCAC	CGCGGATGCT	GGCGCGCACG	TAATCGTATA	GTTCGTGCGA	GGGAGCGAGG
6841	AGGTCGGGAC	CGAGGTTGCT	ACGGGCGGC	TGCTCTGCTC	GGAAGACTAT	CTGCCTGAAG
6901	ATGGCATGTG	AGTTGGATGA	TATGGTTGGA	CGCTGGAAGA	CGTTGAAGCT	GGCGTCTGTG
6961	AGACCTACCG	CGTCACGCAC	GAAGGAGGCG	TAGGAGTCGC	GCAGCTTGTT	GACCAGCTCG
7021	GCGGTGACCT	GCACGTCTAG	GGCGCAGTAG	TCCAGGGTTT	CCTTGATGAT	GTCATACTTA
7081	TCCTGTCCCT	TTTTTTTCCA	CAGCTCGCGG	TTGAGGACAA	ACTCTTCGCG	GTCTTTCCAG
7141	TACTCTTGGA	TCGGAAACCC	GTCGGCCTCC	GAACGGTAAG	AGCCTAGCAT	GTAGAACTGG
7201	TTGACGGCCT	GGTAGGCGCA	GCATCCCTTT	TCTACGGGTA	GCGCGTATGC	CTGCGCGGCC
7261	TTCCGGAGCG	AGGTGTGGGT	GAGCGCAAAG	GTGTCCCTAA	CCATGACTTT	GAGGTACTGG
						CGTGCGCTTT
						TCCCGCGCGA
						GTTAATTACC
						GTAAAGTTCC

FIG. 7C

7561	AAGAAGCGCG	GCATGCCCTT	GATGGAAGGC	ልልጥጥጥጥልል	GTTCCTCGTA	GGTGAGCTCT
	TCAGGGGAGC					
	ACGAATGAGC					
	AACTGGCGAC					
	TCCCAGCGGT					
	TCTCCGCCGA					
	CAAGTATAGG					
	ATCGGGAAGA					
	TAGAAGTCCC					
8101	TGGCAGCGGT	GCACGGGCTG	TACATCCTGC	ACGAGGTTGA	CCTGACGACC	GCGCACAAGG
8161	AAGCAGAGTG	GGAATTTGAG	CCCCTCGCCT	GGCGGGTTTG	GCTGGTGGTC	TTCTACTTCG
8221	GCTGCTTGTC	CTTGACCGTC	TGGCTGCTCG	AGGGGAGTTA	CGGTGGATCG	GACCACCACG
8281	CCGCGCGAGC	CCAAAGTCCA	GATGTCCGCG	CGCGGCGGTC	GGAGCTTGAT	GACAACATCG
8341	CGCAGATGGG	AGCTGTCCAT	GGTCTGGAGC	TCCCGCGGCG	TCAGGTCAGG	CGGGAGCTCC
8401	TGCAGGTTTA	CCTCGCATAG	CCGGGTCAGG	GCGCGGCTA	GGTCCAGGTG	ATACCTGATT
8461	TCCAGGGGCT	GGTTGGTGGC	GGCGTCGATG	GCTTGCAAGA	GGCCGCATCC	CCGCGGCGCG
8521	ACTACGGTAC	CGCGCGGCGG	GCGGTGGGCC	GCGGGGGTGT	CCTTGGATGA	TGCATCTAAA
8581	AGCGGTGACG	CGGGCGGCC	CCCGGAGGTA	GGGGGGCTC	GGGACCCGCC	GGGAGAGGGG
	GCAGGGGCAC				•	
8701	CGAACGCGAC	GACGCGGCGG	TTGATCTCCT	GAATCTGGCG	CCTCTGCGTG	AAGACGACGG
8761	GCCCGGTGAG	CTTGAACCTG	AAAGAGAGTT	CGACAGAATC	AATTTCGGTG	TCGTTGACGG
8821	CGGCCTGGCG	CAAAATCTCC	TGCACGTCTC	CTGAGTTGTC	TTGATAGGCG	ATCTCGGCCA
8881	TGAACTGCTC	GATCTCTTCC	TCCTGGAGAT	CTCCGCGTCC	GGCTCGCTCC	ACGGTGGCGG
	CGAGGTCGTT					
	AGACGCGGCT					
	GATTGAGCTC					
	TGAGGGTGGT					
	ATTCGTTGAT					
	AGTTGAAAAA					
	GCTCGGCGAC					
	CAATCTCCTC					
	GGACACGGCG					
	CGCGGCGACG					
	AGACGCCGCC					
	CGGCGCTAAC					
	GCGAGTCCGC					
	CGCAAGGTAG					
	CGGAGGTGCT					
	GAAGCACCAT					
	CTTCGTTTTG					
	CTTCTTCTTC					
10021	AGTTTGGCCG	TAGGTGGCGC	CCTCTTCCTC	CCATGCGTGT	GACCCCGAAG	CCCCTCATCG

FIG. 7D

10081	GCTGAAGCAG	GGCCAGGTCG	GCGACAACGC	GCTCGGCTAA	TATGGCCTGC	TGCACCTGCG
10141	TGAGGGTAGA	CTGGAAGTCG	TCCATGTCCA	CAAAGCGGTG	GTATGCGCCC	GTGTTGATGG
10201	TGTAAGTGCA	GTTGGCCATA	ACGGACCAGT	TAACGGTCTG	GTGACCCGGC	TGCGAGAGCT
10261	CGGTGTACCT	GAGACGCGAG	TAAGCCCTTG	AGTCAAAGAC	GTAGTCGTTG	CAAGTCCGCA
10321	CCAGGTACTG	GTATCCCACC	AAAAAGTGCG	GCGGCGGCTG	GCGGTAGAGG	GGCCAGCGTA
10381	GGGTGGCCGG	GGCTCCGGGG	GCGAGGTCTT	CCAACATAAG	GCGATGATAT	CCGTAGATGT
10441	ACCTGGACAT	CCAGGTGATG	CCGGCGGCGG	TGGTGGAGGC	GCGCGGAAAG	TCACGGACGC
10501	GGTTCCAGAT	GTTGCGCAGC	GGCAAAAAGT	GCTCCATGGT	CGGGACGCTC	TGGCCGGTCA
10561	GGCGCGCGCA	GTCGTTGACG	CTCTAGACCG	TGCAAAAGGA	GAGCCTGTAA	GCGGGCACTC
10621	TTCCGTGGTC	TGGTGGATAA	ATTCGCAAGG	GTATCATGGC	GGACGACCGG	GGTTCGAACC
10681	CCGGATCCGG	CCGTCCGCCG	TGATCCATGC	GGTTACCGCC	CGCGTGTCGA	ACCCAGGTGT
10741	GCGACGTCAG	ACAACGGGGG	AGCGCTCCTT	TTGGCTTCCT	TCCAGGCGCG	GCGGATGCTG
10801	CGCTAGCTTT	TTTGGCCACT	GGCCGCGCGC	GGCGTAAGCG	GTTAGGCTGG	AAAGCGAAAG
10861	CATTAAGTGG	CTCGCTCCCT	GTAGCCGGAG	GGTTATTTTC	CAAGGGTTGA	GTCGCGGGAC
10921	CCCCGGTTCG	AGTCTCGGGC	CGGCCGGACT	GCGGCGAACG	GGGGTTTGCC	TCCCCGTCAT
10981	GCAAGACCCC	GCTTGCAAAT	TCCTCCGGAA	ACAGGGACGA	GCCCCTTTTT	TGCTTTTCCC
11041	AGATGCATCC	GGTGCTGCGG	CAGATGCGCC	CCCCTCCTCA	GCAGCGGCAA	GAGCAAGAGC
11101	AGCGGCAGAC	ATGCAGGGCA	CCCTCCCCTT	CTCCTACCGC	GTCAGGAGGG	GCAACATCCG
11161	CGGCTGACGC	GGCGGCAGAT	GGTGATTACG	AACCCCCGCG	GCGCCGGACC	CGGCACTACT
11221	TGGACTTGGA	GGAGGGCGAG	GGCCTGGCGC	GGCTAGGAGC	GCCCTCTCCT	GAGCGACACC
11281	CAAGGGTGCA	GCTGAAGCGT	GACACGCGCG	AGGCGTACGT	GCCGCGGCAG	AACCTGTTTC
11341	GCGACCGCGA	GGGAGAGGAG	CCCGAGGAGA	TGCGGGATCG	AAAGTTCCAT	GCAGGGCGCG
11401	AGTTGCGGCA	TGGCCTGAAC	CGCGAGCGGT	TGCTGCGCGA	GGAGGACTTT	GAGCCCGACG
11461	CGCGGACCGG	GATTAGTCCC	GCGCGCGCAC	ACGTGGCGGC	CGCCGACCTG	GTAACCGCGT
11521	ACGAGCAGAC	GGTGAACCAG	GAGATTAACT	TTCAAAAAAG	CTTTAACAAC	CACGTGCGCA
11581	CGCTTGTGGC	GCGCGAGGAG	GTGGCTATAG	GACTGATGCA	TCTGTGGGAC	TTTGTAAGCG
11641	CGCTGGAGCA	AAACCCAAAT	AGCAAGCCGC	TCATGGCGCA	GCTGTTCCTT	ATAGTGCAGC
11701	ACAGCAGGGA	CAACGAGGCA	TTCAGGGATG	CGCTGCTAAA	CATAGTAGAG	CCCGAGGGCC
11761	GCTGGCTGCT	CGATTTGATA	AACATTCTGC	AGAGCATAGT	GGTGCAGGAG	CGCAGCTTGA
11821	GCCTGGCTGA	CAAGGTGGCC	GCCATTAACT	ATTCCATGCT	CAGTCTGGGC	AAGTTTTACG
11881	CCCGCAAGAT	ATACCATACC	CCTTACGTTC	CCATAGACAA	GGAGGTAAAG	ATCGAGGGGT
11941	TCTACATGCG	CATGGCGCTG	AAGGTGCTTA	CCTTGAGCGA	CGACCTGGGC	GTTTATCGCA
12001	ACGAGCGCAT	CCACAAGGCC	GTGAGCGTGA	GCCGGCGGCG	CGAGCTCAGC	GACCGCGAGC
12061	TGATGCACAG	CCTGCAAAGG	GCCCTGGCTG	GCACGGGCAG	CGGCGATAGA	GAGGCCGAGT
12121	CCTACTTTGA	CGCGGGCGCT	GACCTGCGCT	GGGCCCCAAG	CCGACGCGCC	CTGGAGGCAG
12181	CTGGGGCCGG	ACCTGGGCTG	GCGGTGGCAC	CCGCGCGCGC	TGGCAACGTC	GCCGCCGTGG
12241	AGGAATATGA	CGAGGACGAT	GAGTACGAGC	CAGAGGACGG	CGAGTACTA	GCGGTGATGT
12301	TTCTGATCAG	ATGATGCAAG	ACGCAACGGA	CCCGGCGGTG	CGGGCGGCGC	TGCAGAGCCA
12361	GCCGTCCGGC	CTTAACTCCA	CGGACGACTG	GCGCCAGGTC	ATGGACCGCA	TCATGTCGCT
12421	GACTGCGCGC	AACCCTGACG	CGTTCCGGCA	. GCAGCCGCAG	GCCAACCGGC	TCTCCGCAAT
12481	TCTGGAAGCG	GTGGTCCCGG	CGCGCGCAAA	CCCCACGCAC	GAGAAGGTGC	TGGCGATCGT
12541	AAACGCGCTG	GCCGAAAACA	GGGCCATCCG	GCCCGATGAG	GCCGGCCTGC	TCTACGACGC

FIG. 7E

	GCTGCTTCAG					
	GGTGGGGGAT					
	GGGCTCCATG					
	ACAGGAGGAC					
12841	AAGTGAGGTG	TATCAGTCCG	GGCCAGACTA	TTTTTTCCAG	ACCAGTAGAC	AAGGCCTGCA
12901	GACCGTAAAC	CTGAGCCAGG	CTTTCAAGAA	CTTGCAGGGG	CTGTGGGGGG	TGCGGGCTCC
12961	CACAGGCGAC	CGCGCGACCG	TGTCTAGCTT	GCTGACGCCC	AACTCGCGCC	TGTTGCTGCT
13021	GCTAATAGCG	CCCTTCACGG	ACAGTGGCAG	CGTGTCCCGG	GACACATACC	TAGGTCACTT
13081	GCTGACACTG	TACCGCGAGG	CCATAGGTCA	GGCGCATGTG	GACGAGCATA	CTTTCCAGGA
13141	GATTACAAGT	GTTAGCCGCG	CGCTGGGGCA	GGAGGACACG	GGCAGCCTGG	AGGCAACCCT
13201	GAACTACCTG	${\tt CTGACCAACC}$	GGCGGCAAAA	AATCCCCTCG	TTGCACAGTT	TAAACAGCGA
13261	GGAGGAGCGC	ATTTTGCGCT	ATGTGCAGCA	GAGCGTGAGC	CTTAACCTGA	TGCGCGACGG
13321	GGTAACGCCC	AGCGTGGCGC	TGGACATGAC	CGCGCGCAAC	ATGGAACCGG	GCATGTATGC
13381	CTCAAACCGG	CCGTTTATCA	ATCGCCTAAT	GGACTACTTG	CATCGCGCGG	CCGCCGTGAA
13441	CCCCGAGTAT	TTCACCAATG	CCATCTTGAA	CCCGCACTGG	CTACCGCCCC	CTGGTTTCTA
13501	CACCGGGGGA	TTCGAGGTGC	CCGAGGGTAA	CGATGGATTC	CTCTGGGACG	ACATAGACGA
13561	${\tt CAGCGTGTTT}$	TCCCCGCAAC	CGCAGACCCT	GCTAGAGTTG	CAACAACGCG	AGCAGGCAGA
13621	GGCGGCGCTG	CGAAAGGAAA	GCTTCCGCAG	GCCAAGCAGC	TTGTCCGATC	TAGGCGCTGC
13681	GGCCCCGCGG	TCAGATGCTA	GTAGCCCATT	TCCAAGCTTG	ATAGGGTCTC	TTACCAGCAC
13741	TCGCACCACC	CGCCCGCGCC	${\tt TGCTGGGCGA}$	GGAGGAGTAC	CTAAACAACT	CGCTGCTGCA
13801	GCCGCAGCGC	GAAAAGAACC	TGCCTCCGGC	GTTTCCCAAC	AACGGGATAG	AGAGCCTAGT
13861	GGACAAGATG	AGTAGATGGA	AGACGTATGC	GCAGGAGCAC	AGGGATGTGC	CCGGCCCGCG
13921	CCCGCCCACC	CGTCGTCAAA	GGCACGACCG	TCAGCGGGGT	CTGGTGTGGG	AGGACGATGA
13981	CTCGGCAGAC	GACAGCAGCG	${\tt TCTTGGATTT}$	GGGAGGGAGT	GGCAACCCGT	TTGCACACCT
14041	TCGCCCCAGG	CTGGGGAGAA	TGTTTTAAAA	AAAGCATGAT	GCAAAATAAA	AAACTCACCA
14101	AGGCCATGGC	ACCGAGCGTT	${\tt GGTTTTCTTG}$	${\tt TATTCCCCTT}$	AGTATGCGGC	GCGCGGCGAT
14161	GTATGAGGAA	GGTCCTCCTC	CCTCCTACGA	GAGCGTGGTG	AGCGCGGCGC	CAGTGGCGGC
14221	GGCGCTGGGT	TCACCCTTCG	ATGCTCCCCT	GGACCCGCCG	TTCGTGCCTC	CGCGGTACCT
14281	GCGGCCTACC	GGGGGGAGAA	ACAGCATCCG	TTACTCTGAG	TTGGCACCCC	TATTCGACAC
14341	CACCCGTGTG	TACCTTGTGG	ACAACAAGTC	AACGGATGTG	GCATCCCTGA	ACTACCAGAA
14401	CGACCACAGC	AACTTTCTAA	CCACGGTCAT	TCAAAACAAT	GACTACAGCC	CGGGGGAGGC
14461	AAGCACACAG	ACCATCAATC	TTGACGACCG	GTCGCACTGG	GGCGGCGACC	TGAAAACCAT
14521	CCTGCATACC	AACATGCCAA	ATGTGAACGA	GTTCATGTTT	ACCAATAAGT	TTAAGGCGCG
14581	GGTGATGGTG	TCGCGCTCGC	TTACTAAGGA	CAAACAGGTG	GAGCTGAAAT	ACGAGTGGGT
14641	GGAGTTCACG	CTGCCCGAGG	GCAACTACTC	CGAGACCATG	ACCATAGACC	TTATGAACAA
14701	CGCGATCGTG	GAGCACTACT	TGAAAGTGGG	CAGGCAGAAC	GGGGTTCTGG	AAAGCGACAT
14761	CGGGGTAAAG	TTTGACACCC	GCAACTTCAG	ACTGGGGTTT	GACCCAGTCA	CTGGTCTTGT
14821	CATGCCTGGG	GTATATACAA	ACGAAGCCTT	CCATCCAGAC	ATCATTTTGC	TGCCAGGATG
14881	CGGGGTGGAC	TTCACCCACA	GCCGCCTGAG	CAACTTGTTG	GGCATCCGCA	AGCGGCAACC
14941	CTTCCAGGAG	GGCTTTAGGA	TCACCTACGA	TGACCTGGAG	GGTGGTAACA	TTCCCGCACT
15001	GTTGGATGTG	GACGCCTACC	AGGCAAGCTT	GAAAGATGAC	ACCGAACAGG	GCGGGGGTGG
15061	CGCAGGCGGC	GGCAACAACA	GTGGCAGCGG	CGCGGAAGAG	AACTCCAACG	CGGCAGCTGC

FIG. 7F

					CGCGGCGACA	
					GAAGCTGCCG	
					CCGGTGATTA	
					GACAGCACCT	
					GCCGGGATCC	
15421	CCTGCTTTGC	ACTCCTGACG	TAACCTGCGG	CTCGGAGCAG	GTATACTGGT	CGTTGCCCGA
15481	CATGATGCAA	GACCCCGTGA	CCTTCCGCTC	CACGCGCCAG	ATCAGCAACT	TTCCGGTGGT
15541	GGGCGCCGAG	CTGTTGCCCG	TGCACTCCAA	GAGCTTCTAC	AACGACCAGG	CCGTCTACTC
15601	CCAGCTCATC	CGCCAGTTTA	CCTCTCTGAC	CCACGTGTTC	AATCGCTTTC	CCGAGAACCA
15661	GATTTTGGCG	CGCCCGCCAG	CCCCACCAT	CACCACCGTC	AGTGAAAACG	TTCCTGCTCT
15721	CACAGATCAC	GGGACGCTAC	CGCTGCGCAA	CAGCATCGGA	GGAGTCCAGC	GAGTGACCAT
15781	TACTGACGCC	AGACGCCGCA	${\tt CCTGCCCTA}$	CGTTTACAAG	GCCCTGGGCA	TAGTCTCGCC
15841	GCGCGTCCTA	TCGAGCCGCA	${\tt CTTTTTGAGC}$	AAGCATGTCC	ATCCTTATAT	CGCCCAGCAA
					GGCGGGGCCA	
					CCCTGGGGCG	
16021	CGGCCGCACT	GGGCGCACCA	CCGTCGATGA	CGCCATCGAC	GCGGTGGTGG	AGGAGGCGCG
16081	CAACTACACG	CCCACGCCGC	CGCCAGTGTC	CACCGTGGAC	GCGGCCATTC	AGACCGTGGT
16141	GCGCGGAGCC	CGGCGCTACG	CTAAAATGAA	GAGACGGCGG	AGGCGCGTAG	CACGTCGCCA
					GCCCTGCTTA	
16261	TCGCACCGGC	CGACGGGCGG	CCATGCGAGC	CGCTCGAAGG	CTGGCCGCGG	GTATTGTCAC
16321	TGTGCCCCCC	AGGTCCAGGC	GACGAGCGGC	CGCCGCAGCA	GCCGCGGCCA	TTAGTGCTAT
					TCGGTTAGCG	
					AAAAACTACT	
					ATGTCCAAGC	
					CCCCGAAGA	
					AAGAAAGATG	
					CCCAGGCGAC	
					ACCGTAGTCT	
					GTGTACGGCG	
					GGAAAGCGGC	
					CTAAAGCCCG	
					GGCCTAAAGC	
17041	TGACTTGGCA	CCCACCGTGC	AGCTGATGGT	ACCCAAGCGT	CAGCGACTGG	AAGATGTCTT
17101	GGAAAAAATG	ACCGTGGAGC	CTGGGCTGGA	GCCCGAGGTC	CGCGTGCGGC	CAATCAAGCA
					ATACCCACCA	
						CCTCGGCGGT
						CGGAGGTGCA
						CAAGGAAGTA
					CCTTCCATCG	
17461	CGGCTATCGT	GGCTACACCT	ACCGCCCCAG	AAGACGAGCA	ACTACCCGAC	GCCGAACCAC
						TTTCCGTGCG
17581	CAGGGTGGCT	CGCGAAGGAG	GCAGGACCCT	GGTGCTGCCA	ACAGCGCGCT	ACCACCCCAG

FIG. 7G

	CATCGTTTAA					
	TTTCCCGGTG					
	CCTGACGGGC					
	GCGCGGCGGT					
	CGGAATTGCA					
	GAAAAATCAA					
	AATGGAAGAC					
18061	AAACTGGCAA	GATATCGGCA	CCAGCAATAT	GAGCGGTGGC	GCCTTCAGCT	GGGGCTCGCT
18121	GTGGAGCGGC	${\tt ATTAAAAATT}$	TCGGTTCCGC	CGTTAAGAAC	TATGGCAGCA	AAGCCTGGAA
18181	CAGCAGCACA	GGCCAGATGC	TGAGGGACAA	GTTGAAAGAG	CAAAATTTCC	AACAAAAGGT
18241	GGTAGATGGC	CTGGCCTCTG	GCATTAGCGG	GGTGGTGGAC	CTGGCCAACC	AGGCAGTGCA
18301	AAATAAGATT	AACAGTAAGC	TTGATCCCCG	CCCTCCCGTA	GAGGAGCCTC	CACCGGCCGT
18361	${\tt GGAGACAGTG}$	TCTCCAGAGG	GGCGTGGCGA	AAAGCGTCCG	CGACCCGACA	GGGAAGAAAC
18421	${\tt TCTGGTGACG}$	CAAATAGACG	AGCCTCCCTC	GTACGAGGAG	GCACTAAAGC	AAGGCCTGCC
18481	CACCACCCGT	CCCATCGCGC	CCATGGCTAC	CGGAGTGCTG	GGCCAGCACA	CACCCGTAAC
18541	GCTGGACCTG	CCTCCCCCG	CCGACACCCA	GCAGAAACCT	GTGCTGCCAG	GCCCGTCCGC
18601	CGTTGTTGTA	ACCCGTCCTA	GCCGCGCGTC	CCTGCGCCGC	GCCGCCAGCG	GTCCGCGATC
18661	${\tt GTTGCGGCCC}$	GTAGCCAGTG	GCAACTGGCA	AAGCACACTG	AACAGCATCG	TGGGTTTGGG
18721	${\tt GGTGCAATCC}$	CTGAAGCGCC	GACGATGCTT	CTGATAGCTA	ACGTGTCGTA	TGTGTGTCAT
18781	GTATGCGTCC	ATGTCGCCGC	CAGAGGAGCT	GCTGAGCCGC	CGCGCGCCCG	CTTTCCAAGA
18841	TGGCTACCCC	TTCGATGATG	CCGCAGTGGT	CTTACATGCA	CATCTCGGGC	CAGGACGCCT
18901	CGGAGTACCT	GAGCCCCGGG	CTGGTGCAGT	TCGCCCGCGC	CACCGAGACG	TACTTCAGCC
18961	TGAATAACAA	GTTTAGAAAC	CCCACGGTGG	CGCCTACGCA	CGACGTGACC	ACAGACCGGT
19021	CTCAGCGTTT	GACGCTGCGG	TTCATCCCCG	TGGACCGCGA	GGATACTGCG	TACTCGTACA
19081	AGGCGCGGTT	CACCCTAGCT	${\tt GTGGGTGATA}$	ACCGTGTGCT	AGACATGGCT	TCCACGTACT
19141	TTGACATCCG	CGGCGTGCTG	GACAGGGGCC	CTACTTTTAA	GCCCTACTCT	GGCACTGCCT
19201	ACAACGCACT	GGCCCCCAAG	GGTGCCCCCA	ACTCGTGCGA	GTGGGAACAA	AATGAAACTG
19261	CACAAGTGGA	TGCTCAAGAA	CTTGACGAAG	AGGAGAATGA	AGCCAATGAA	GCTCAGGCGC
19321	GAGAACAGGA	ACAAGCTAAG	AAAACCCATG	TATATGCCCA	GGCTCCACTG	TCCGGAATAA
19381	AAATAACTAA	AGAAGGTCTA	CAAATAGGAA	CTGCCGACGC	CACAGTAGCA	GGTGCCGGCA
19441	AAGAAATTTT	CGCAGACAAA	ACTTTTCAAC	CTGAACCACA	AGTAGGAGAA	TCTCAATGGA
	ACGAAGCGGA					
19561	CCTGCTATGG	CTCATACGCT	AGACCCACCA	ATTCCAACGG	CGGACAGGGC	GTTATGGTTG
19621	AACAAAATGG	TAAATTGGAA	AGTCAAGTCG	AAATGCAATT	TTTTTCCACA	TCCACAAATG
19681	CCACAAATGA	AGTTAACAAT	ATACAACCAA	CAGTTGTATT	GTACAGCGAA	GATGTAAACA
19741	TGGAAACTCC	AGATACTCAT	СТТТСТТАТА	AACCTAAAAT	GGGGGATAAA	AATGCCAAAG
19801	TCATGCTTGG	ACAACAAGCA	ATGCCAAACA	GACCAAATTA	CATTGCTTTT	AGAGACAATT
19861	TTATTGGTCT	CATGTATTAC	AACAGCACAG	GTAACATGGG	TGTCCTTGCT	GGTCAGGCAT
19921	CGCAGTTGAA	CGCTGTTGTA	GATTTGCAAG	ACAGAAACAC	AGAGCTGTCC	TACCAGCTTT
19981	TGCTTGATTC	AATTGGCGAC	AGAACAAGAT	ACTTTTCAAT	GTGGAATCAA	GCTGTTGACA
20041	GCTATGATCC	AGATGTCAGA	ATTATTGAGA	ACCATGGAAC	TGAGGATGAG	TTGCCAAATT
20101	ATTGCTTTCC	TCTTGGTGGA	ATTGGGATTA	CTGACACTTT	TCAAGCTGTT	AAAACAACTG

FIG. 7H

PCT/US02/32512

20161	CTGCTAACGG	GGACCAAGGC	AATACTACCT	GGCAAAAAGA	TTCAACATTT	GCAGAACGCA
20221	ATGAAATAGG	GGTGGGAAAT	AACTTTGCCA	TGGAAATTAA	CCTGAATGCC	AACCTATGGA
20281	GAAATTTCCT	TTACTCCAAT	ATTGCGCTGT	ACCTGCCAGA	CAAGCTAAAA	TACAACCCCA
20341	CCAATGTGGA	AATATCTGAC	AACCCCAACA	CCTACGACTA	CATGAACAAG	CGAGTGGTGG
20401	CTCCTGGGCT	TGTAGACTGC	TACATTAACC	TTGGGGCGCG	CTGGTCTCTG	GACTACATGG
20461	ACAACGTTAA	TCCCTTTAAC	CACCACCGCA	ATGCGGGCCT	GCGTTACCGC	TCCATGTTGT
20521	TGGGAAACGG	CCGCTACGTG	CCCTTTCACA	TTCAGGTGCC	CCAAAAGTTT	TTTGCCATTA
20581	AAAACCTCCT	CCTCCTGCCA	GGCTCATACA	CATATGAATG	GAACTTCAGG	AAGGATGTTA
20641	ACATGGTTCT	GCAGAGCTCT	CTGGGAAACG	ACCTTAGAGT	TGACGGGGCT	AGCATTAAGT
20701	TTGACAGCAT	TTGTCTTTAC	GCCACCTTCT	TCCCCATGGC	CCACAACACG	GCCTCCACGC
		GCTCAGAAAT				
		ATATCCCATA				
		AGCATTTCGC				
		AGGCTACGAC				
		TCTTAATCAC				
		CAACGACCGC				
		CTATAACGTA				
		CTACAATATT				
		GTTCTTCAGA				
		TTATCAGCAG				
		TCCCACCATG				
		AACCGCGGTT				
		CCCCTTCTCC				
		CTACGCAAAC				
		CACCCTTCTT				
		CGGCGTCATC				
		AAGAAGCAAG				
		GCCATTGTCA				
		CCAGGCTTTG				
		ACTGGGGGCG				
		TTTGAGCCCT				
22021	TGAGTACGAG	TCACTCCTGC	GCCGTAGCGC	CATTGCCTCT	TCCCCCGACC	GCTGTATAAC
22081	GCTGGAAAAG	TCCACCCAAA	GCGTGCAGGG	GCCCAACTCG	GCCGCCTGTG	GCCTATTCTG
22141	CTGCATGTTT	CTCCACGCCT	TTGCCAACTG	GCCCCAAACT	CCCATGGATC	ACAACCCCAC
22201	CATGAACCTT	ATTACCGGGG	TACCCAACTC	CATGCTTAAC	AGTCCCCAGG	TACAGCCCAC
						CCTACTTCCG
						ACATGTAAAA
22381	ATAATGTACT	AGGAGACACT	TTCAATAAAG	GCAAATGTTT	TTATTTGTAC	ACTCTCGGGT
22441	GATTATTTAC	CCCCACCCTT	GCCGTCTGCG	CCGTTTAAAA	ATCAAAGGGG	TTCTGCCGCG
22501	CATCGCTATG	CGCCACTGGC	AGGGACACGT	TGCGATACTG	GTGTTTAGTG	CTCCACTTAA
22561	ACTCAGGCAC	AACCATCCGC	GGCAGCTCGG	TGAAGTTTTC	ACTCCACAGG	CTGCGCACCA
22621	TCACCAACGO	GTTTAGCAGG	TCGGGCGCCG	ATATCTTGAA	GTCGCAGTTG	GGGCCTCCGC

FIG. 71

22681	CCTGCGCGCG	CGAGTTGCGA	TACACAGGGT	TACAGCACTG	GAACACTATC	AGCGCCGGGT
22741	GGTGCACGCT	GGCCAGCACG	CTCTTGTCGG	AGATCAGATC	CGCGTCCAGG	TCCTCCGCGT
22801	TGCTCAGGGC	GAACGGAGTC	AACTTTGGTA	GCTGCCTTCC	CAAAAAGGGT	GCATGCCCAG
22861	GCTTTGAGTT	GCACTCGCAC	CGTAGTGGCA	TCAGAAGGTG	ACCGTGCCCA	GTCTGGGCGT
22921	TAGGATACAG	CGCCTGCATG	AAAGCCTTGA	TCTGCTTAAA	AGCCACCTGA	GCCTTTGCGC
22981	CTTCAGAGAA	GAACATGCCG	CAAGACTTGC	CGGAAAACTG	ATTGGCCGGA	CAGGCCGCGT
23041	CATGCACGCA	GCACCTTGCG	TCGGTGTTGG	AGATCTGCAC	CACATTTCGG	CCCCACCGGT
23101	TCTTCACGAT	CTTGGCCTTG	CTAGACTGCT	CCTTCAGCGC	GCGCTGCCCG	TTTTCGCTCG
23161	TCACATCCAT	TTCAATCACG	TGCTCCTTAT	TTATCATAAT	GCTCCCGTGT	AGACACTTAA
23221	GCTCGCCTTC	GATCTCAGCG	CAGCGGTGCA	GCCACAACGC	GCAGCCCGTG	GGCTCGTGGT
23281	GCTTGTAGGT	TACCTCTGCA	AACGACTGCA	GGTACGCCTG	CAGGAATCGC	CCCATCATCG
23341	TCACAAAGGT	CTTGTTGCTG	GTGAAGGTCA	GCTGCAACCC	GCGGTGCTCC	TCGTTTAGCC
23401	AGGTCTTGCA	TACGGCCGCC	AGAGCTTCCA	CTTGGTCAGG	CAGTAGCTTG	AAGTTTGCCT
23461	TTAGATCGTT	ATCCACGTGG	TACTTGTCCA	TCAACGCGCG	CGCAGCCTCC	ATGCCCTTCT
23521	CCCACGCAGA	CACGATCGGC	AGGCTCAGCG	GGTTTATCAC	CGTGCTTTCA	CTTTCCGCTT
23581	CACTGGACTC	TTCCTTTTCC	TCTTGCATCC	GCATACCCCG	CGCCACTGGG	TCGTCTTCAT
23641	TCAGCCGCCG	CACCGTGCGC	${\tt TTACCTCCCT}$	${\tt TGCCGTGCTT}$	GATTAGCACC	GGTGGGTTGC
23701	TGAAACCCAC	CATTTGTAGC	GCCACATCTT	${\tt CTCTTTCTTC}$	CTCGCTGTCC	ACGATCACCT
23761	CTGGGGATGG	CGGGCGCTCG	GGCTTGGGAG	${\tt AGGGGCGCTT}$	${\tt CTTTTTCTTT}$	TTGGACGCAA
23821	TGGCCAAATC	CGCCGTCGAG	GTCGATGGCC	GCGGGCTGGG	TGTGCGCGGC	ACCAGCGCAT
23881	CTTGTGACGA	GTCTTCTTCG	TCCTCGGACT	CGAGACGCCG	CCTCAGCCGC	TTTTTTGGGG
23941	GCGCGCGGG	AGGCGGCGGC	GACGGCGACG	GGGACGAGAC	GTCCTCCATG	GTTGGTGGAC
24001	GTCGCGCCGC	ACCGCGTCCG	CGCTCGGGGG	TGGTTTCGCG	CTGCTCCTCT	TCCCGACTGG
24061	CCATTTCCTT	CTCCTATAGG	CAGAAAAAGA	TCATGGAGTC	AGTCGAGAAG	GAGGACAGCC
24121	TAACCGCCCC	CTTTGAGTTC	GCCACCACCG	CCTCCACCGA	TGCCGCCAAC	GCGCCTACCA
24181	CCTTCCCCGT	CGAGGCACCC	CCGCTTGAGG	AGGAGGAAGT	GATTATCGAG	CAGGACCCAG
24241	GTTTTGTAAG	CGAAGACGAC	GAAGATCGCT	CAGTACCAAC	AGAGGATAAA	AAGCAAGACC
24301	AGGACGACGC	AGAGGCAAAC	GAGGAACAAG	TCGGGCGGG	GGACCAAAGG	CATGGCGACT
24361	ACCTAGATGT	GGGAGACGAC	GTGCTGTTGA	AGCATCTGCA	GCGCCAGTGC	GCCATTATCT
24421	GCGACGCGTT	GCAAGAGCGC	AGCGATGTGC	CCCTCGCCAT	AGCGGATGTC	AGCCTTGCCT
24481	ACGAACGCCA	CCTGTTCTCA	CCGCGCGTAC	CCCCCAAACG	CCAAGAAAAC	GGCACATGCG
24541	AGCCCAACCC	GCGCCTCAAC	TTCTACCCCG	TATTTGCCGT	GCCAGAGGTG	CTTGCCACCT
24601	ATCACATCTT	TTTCCAAAAC	TGCAAGATAC	CCCTATCCTG	CCGTGCCAAC	CGCAGCCGAG
24661	CGGACAAGCA	GCTGGCCTTG	CGGCAGGGCG	CTGTCATACC	TGATATCGCC	TCGCTCGACG
24721	AAGTGCCAAA	AATCTTTGAG	GGTCTTGGAC	GCGACGAGAA	GCGCGCGCA	AACGCTCTGC
24781	AACAAGAAAA	CAGCGAAAAT	GAAAGTCACT	GTGGAGTGCT	GGTGGAACTT	GAGGGTGACA
24841	ACGCGCGCCT	AGCCGTGCTG	AAACGCAGCA	TCGAGGTCAC	CCACTTTGCC	TACCCGGCAC
24901	TTAACCTACC	CCCCAAGGTT	ATGAGCACAG	TCATGAGCGA	GCTGATCGTG	CGCCGTGCAC
24961	GACCCTGGA	GAGGGATGCA	AACTTGCAAG	AACAAACCGA	GGAGGGCCTA	CCCGCAGTTG
25021	GCGATGAGCA	GCTGGCGCGC	TGGCTTGAGA	CGCGCGAGCC	TGCCGACTTG	GAGGAGCGAC
25081	GCAAGCTAAT	GATGGCCGCA	GTGCTTGTTA	CCGTGGAGCT	TGAGTGCATG	CAGCGGTTCT
25141	TTGCTGACCC	GGAGATGCAG	CGCAAGCTAG	AGGAAACGTT	GCACTACACC	TTTCGCCAGG

FIG. 7J

25201	GCTACGTGCG	CCAGGCCTGC	AAAATTTCCA	ACGTGGAGCT	CTGCAACCTG	GTCTCCTACC
25261	TTGGAATTTT	GCACGAAAAC	CGCCTTGGGC	AAAACGTGCT	TCATTCCACG	CTCAAGGGCG
25321	AGGCGCGCCG	CGACTACGTC	CGCGACTGCG	TTTACTTATT	TCTGTGCTAC	ACCTGGCAAA
25381	CGGCCATGGG	CGTGTGGCAG	CAGTGCCTGG	AGGAGCGCAA	CCTGAAGGAG	CTGCAGAAGC
		AAACTTGAAG				
		CATTATCTTC				
25561	ACTTCACCAG	TCAAAGCATG	TTGCAAAACT	TTAGGAACTT	TATCCTAGAG	CGTTCAGGAA
25621	TTCTGCCCGC	CACCTGCTGT	GCGCTTCCTA	GCGACTTTGT	GCCCATTAAG	TACCGTGAAT
25681	GCCTCCGCC	GCTTTGGGGT	CACTGCTACC	TTCTGCAGCT	AGCCAACTAC	CTTGCCTACC
25741	ACTCCGACAT	CATGGAAGAC	GTGAGCGGTG	ACGGCCTACT	GGAGTGTCAC	TGTCGCTGCA
25801	ACCTATGCAC	CCCGCACCGC	TCCCTGGTCT	GCAATTCACA	ACTGCTTAGC	GAAAGTCAAA
25861	TTATCGGTAC	CTTTGAGCTG	CAGGGTCCCT	CGCCTGACGA	AAAGTCCGCG	GCTCCGGGGT
		TCCGGGGCTG				
		CGAGATTAGG				
26041	CCGCCTGCGT	CATTACCCAG	GGCCACATCC	TTGGCCAATT	GCAAGCCATT	AACAAAGCCC
26101	GCCAAGAGTT	TCTGCTACGA	AAGGGACGGG	GGGTTTACTT	GGACCCCCAG	TCCGGCGAGG
26161	AGCTCAACCC	AATCCCCCCG	CCGCCGCAGC	CCTATCAGCA	GCCGCGGGCC	CTTGCTTCCC
26221	AGGATGGCAC	CCAAAAAGAA	GCTGCAGCTG	CCGCCGCCGC	CACCCACGGA	CGAGGAGGAA
26281	TACTGGGACA	GTCAGGCAGA	GGAGGTTTTG	GACGAGGAGG	AGGAGATGAT	GGAAGACTGG
26341	GACAGCCTAG	ACGAGGAAGC	TTCCGAGGCC	GAAGAGGTGT	CAGACGAAAC	ACCGTCACCC
26401	TCGGTCGCAT	TCCCCTCGCC	GGCGCCCCAG	AAATCGGCAA	CCGTTCCCAG	CATTGCTACA
26461	ACCTCCGCTC	CTCAGGCGCC	GCCGGCACTG	CCCGTTCGCC	GACCCAACCG	TAGATGGGAC
26521	ACCACTGGAA	CCAGGGCCGG	TAAGTCTAAG	CAGCCGCCGC	CGTTAGCCCA	AGAGCAACAA
26581	CAGCGCCAAG	GCTACCGCTC	GTGGCGCGTG	CACAAGAACG	CCATAGTTGC	TTGCTTGCAA
26641	GACTGTGGGG	GCAACATCTC	CTTCGCCCGC	CGCTTTCTTC	TCTACCATCA	CGGCGTGGCC
26701	TTCCCCCGTA	ACATCCTGCA	TTACTACCGT	CATCTCTACA	GCCCCTACTG	CACCGGCGGC
26761	AGCGGCAGCA	ACAGCAGCGG	CCACGCAGAA	GCAAAGGCGA	CCGGATAGCA	AGACTCTGAC
26821	AAAGCCCAAG	AAATCCACAG	CGGCGGCAGC	AGCAGGAGGA	GGAGCACTGC	GTCTGGCGCC
26881	CAACGAACCC	GTATCGACCC	GCGAGCTTAG	AAACAGGATT	TTTCCCACTC	TGTATGCTAT
26941	ATTTCAACAG	AGCAGGGGCC	AAGAACAAGA	GCTGAAAATA	AAAAACAGGT	CTCTGCGCTC
						CGCTGGAAGA
27061	CGCGGAGGCT	CTCTTCAGCA	AATACTGCGC	GCTGACTCTT	AAGGACTAGT	TTCGCGCCCT
27121	TTCTCAAATT	TAAGCGCGAA	AACTACGTCA	TCTCCAGCGG	CCACACCCGG	CGCCAGCACC
27181	TGTCGTCAGC	GCCATTATGA	GCAAGGAAAT	TCCCACGCCC	TACATGTGGA	GTTACCAGCC
27241	ACAAATGGGA	CTTGCGGCTG	GAGCTGCCCA	AGACTACTCA	ACCCGAATAA	ACTACATGAG
27301	CGCGGGACCC	CACATGATAT	CCCGGGTCAA	CGGAATCCGC	GCCCACCGAA	ACCGAATTCT
27361	CCTCGAACAG	GCGGCTATTA	CCACCACACC	TCGTAATAAC	CTTAATCCCC	GTAGTTGGCC
27421	CGCTGCCCTG	GTGTACCAGG	AAAGTCCCGC	TCCCACCACT	GTGGTACTTC	CCAGAGACGC
27481	CCAGGCCGAA	GTTCAGATGA	CTAACTCAGG	GGCGCAGCTT	GCGGGCGGCT	TTCGTCACAG
27541	GGTGCGGTCG	CCCGGGCAGG	GTATAACTCA	CCTGAAAATC	AGAGGGCGAG	GTATTCAGCT
27601	CAACGACGAG	TCGGTGAGCT	CCTCTCTTGG	TCTCCGTCCG	GACGGGACAT	TTCAGATCGG
27661	CGGCGCTGGC	CGCTCTTCAT	TTACGCCCCG	TCAGGCGATC	CTAACTCTGC	AGACCTCGTC

FIG. 7K

					ATTGAGGAGT	
					CCGGACCAGT	
					ATGACCAGTG	
27901	GCAACTGCGC	CTGACACACC	TCGACCACTG	CCGCCGCCAC	AAGTGCTTTG	CCCGCGGCTC
27961	CGGTGAGTTT	TGTTACTTTG	AATTGCCCGA	AGAGCATATC	GAGGGCCCGG	CGCACGGCGT
28021	CCGGCTCACC	ACCCAGGTAG	AGCTTACACG	TAGCCTGATT	CGGGAGTTTA	CCAAGCGCCC
28081	CCTGCTAGTG	GAGCGGGAGC	GGGGTCCCTG	TGTTCTGACC	GTGGTTTGCA	ACTGTCCTAA
28141	CCCTGGATTA	CATCAAGATC	TTTGTTGTCA	TCTCTGTGCT	GAGTATAATA	AATACAGAAA
28201	TTAGAATCTA	CTGGGGCTCC	TGTCGCCATC	CTGTGAACGC	CACCGTTTTT	ACCCACCCAA
28261	AGCAGACCAA	AGCAAACCTC	ACCTCCGGTT	TGCACAAGCG	GGCCAATAAG	TACCTTACCT
28321	GGTACTTTAA	CGGCTCTTCA	TTTGTAATTT	ACAACAGTTT	CCAGCGAGAC	GAAGTAAGTT
28381	TGCCACACAA	CCTTCTCGGC	TTCAACTACA	CCGTCAAGAA	AAACACCACC	ACCACCCTCC
28441	${\tt TCACCTGCCG}$	GGAACGTACG	AGTGCGTCAC	CGGTTGCTGC	GCCCACACCT	ACAGCCTGAG
28501	CGTAACCAGA	CATTACTCCC	ATTTTCCCAA	AACAGGAGGT	GAGCTCAACT	CCCGGAACTC
28561	AGGTCAAAAA	AGCATTTTGC	GGGGTGCTGG	GATTTTTTAA	TTAAGTATAT	GAGCAATTCA
28621	AGTAACTCTA	CAAGCTTGTC	TAATTTTTCT	${\tt GGAATTGGGG}$	TCGGGGTTAT	CCTTACTCTT
28681	GTAATTCTGT	TTATTCTTAT	ACTAGCACTT	CTGTGCCTTA	GGGTTGCCGC	CTGCTGCACG
28741	CACGTTTGTA	CCTATTGTCA	${\tt GCTTTTTAAA}$	CGCTGGGGGC	GACATCCAAG	ATGAGGTACA
28801	${\tt TGATTTTAGG}$	CTTGCTCGCC	${\tt CTTGCGGCAG}$	TCTGCAGCGC	TGCCAAAAAG	GTTGAGTTTA
28861	AGGAACCAGC	${\tt TTGCAATGTT}$	ACATTTAAAT	CAGAAGCTAA	TGAATGCACT	ACTCTTATAA
28921	AATGCACCAC	AGAACATGAA	AAGCTTATTA	TTCGCCACAA	AGACAAAATT	GGCAAGTATG
28981	CTGTATATGC	TATTTGGCAG	CCAGGTGACA	CTAACGACTA	TAATGTCACA	GTCTTCCAAG
29041	GTGAAAATCG	${\tt TAAAACTTTT}$	ATGTATAAAT	TTCCATTTTA	TGAAATGTGC	GATATTACCA
29101	TGTACATGAG	CAAACAGTAC	AAGTTGTGGC	CCCCACAAAA	GTGTTTAGAG	AACACTGGCA
29161	CCTTTTGTTC	CACCGCTCTG	CTTATTACAG	CGCTTGCTTT	GGTATGTACC	TTACTTTATC
29221	TCAAATACAA	AAGCAGACGC	AGTTTTATTG	ATGAAAAGAA	AATGCCTTGA	TTTTCCGCTT
29281	GCTTGTATTC	CCCTGGACAA	TTTACTCTAT	GTGGGATATG	CGCCAGGCGG	GAAAGATTAT
29341	ACCCACAACC	TTCAAATCAA	ACTTTCCTGG	ACGTTAGCGC	CTGACTTCTG	CCAGCGCCTG
29401	CACTGCAAAT	TTGATCAAAC	CCAGCTTCAG	CTTGCCTGCT	CCAGAGATGA	CCGGCTCAAC
29461	CATCGCGCCC	ACAACGGACT	ATCGCAACAC	CACTGCTACC	GGACTAAAAT	CTGCCCTAAA
29521	TTTACCCCAA	GTTCATGCCT	TTGTCAATGA	CTGGGCGAGC	TTGGGCATGT	GGTGGTTTTC
29581	CATAGCGCTT	ATGTTTGTTT	GCCTTATTAT	TATGTGGCTT	ATTTGTTGCC	TAAAGCGCAG
29641	ACGCGCCAGA	CCCCCATCT	ATAGGCCTAT	CATTGTGCTC	AACCCACACA	ATGAAAAAAT
29701	TCATAGATTG	GACGGTCTCA	AACCATGTTC	TCTTCTTTTA	CAGTATGATT	AAATGAGACA
29761	TGATTCCTCG	AGTCCTTATA	TTATTGACCC	TTGTTGCGCT	TTTCTGTGCG	TGCTCTACAT
29821	TGGCTGCGGT	CGCTCACATC	GAAGTAGATT	GCATCCCACC	TTTCACAGTT	TACCTGCTTT
29881	ACGGATTTGT	CACCCTTATC	CTCATCTGCA	GCCTCGTCAC	TGTAGTCATC	GCCTTCATTC
29941	AGTTCATTGA	CTGGATTTGT	GTGCGCATTG	CGTACCTTAG	GCACCATCCG	CAATACAGAG
30001	ACAGGACTAT	AGCTGATCTT	CTCAGAATTC	TTTAATTATG	AAACGGATTG	TCACTTTTGT
30061	TTTGCTGATT	TTCTGCGCCC	TACCTGTGCT	TTGCTCCCAA	ACCTCAGCGC	CTCCCAAAAG
30121	ACATATTTCC	TGCAGATTCA	CTCAAATATG	GAACATTCCC	AGCTGCTACA	ACAAACAGAG
30181	CGATTTGTCA	GAAGCCTGGT	TATACGCCAT	CATCTCTGTC	ATGGTTTTTT	GCAGTACCAT

FIG. 7L

PCT/US02/32512

30241	TTTTGCCCTA	GCCATATACC	CATACCTTGA	CATTGGTTGG	AATGCCATAG	ATGCCATGAA
30301	CCACCCTACT	TTCCCAGCGC	CCAATGTCAT	ACCACTGCAA	CAGGTTATTG	CCCCAATCAA
30361	TCAGCCTCGC	CCCCCTTCTC	CCACCCCCAC	TGAGATTAGC	TACTTTAATT	TGACAGGTGG
30421	AGATGACTGA	ATCTCTAGAT	CTAGAATTGG	ATGGAATTAA	CACCGAACAG	CGCCTACTAG
30481	AAAGGCGCAA	GGCGGCGTCC	GAGCGAGAAC	GCCTAAAACA	AGAAGTTGAA	GACATGGTTA
30541	ACCTGCACCA	GTGTAAAAGA	GGTATCTTTT	GTGTGGTCAA	GCAGGCCAAA	CTTACCTACG
30601	AAAAAACCAC	TACCGGCAAC	CGCCTTAGCT	ACAAGCTACC	CACCCAGCGC	CAAAAACTGG
30661	TGCTTATGGT	GGGAGAAAA	CCTATCACCG	TCACCCAGCA	CTCGGCAGAA	ACAGAAGGCT
30721	GCCTGCACTT	CCCCTATCAG	GGTCCAGAGG	ACCTCTGCAC	TCTTATTAAA	ACCATGTGTG
30781	GCATTAGAGA	TCTTATTCCA	TTCAACTAAC	ААТАААСАСА	CAATAAATTA	CTTACTTAAA
30841	ATCAGTCAGC	AAATCTTTGT	CCAGCTTATT	CAGCATCACC	TCCTTTCCCT	CCTCCCAACT
30901	CTGGTATTTC	AGCAGCCTTT	TAGCTGCGAA	CTTTCTCCAA	AGTCTAAATG	GGATGTCAAA
30961	TTCCTCATGT	TCTTGTCCCT	CCGCACCCAC	TATCTTCATA	TTGTTGCAGA	TGAAACGCGC
31021	CAGACCGTCT	GAAGACACCT	TCAACCCTGT	GTACCCATAT	GACACGGAAA	CCGGCCCTCC
31081	AACTGTGCCT	TTCCTTACCC	CTCCCTTTGT	GTCGCCAAAT	GGGTTCCAAG	AAAGTCCCCC
31141	CGGAGTGCTT	TCTTTGCGTC	TTTCAGAACC	TTTGGTTACC	TCACACGGCA	TGCTTGCGCT
31201	AAAAATGGGC	AGCGGCCTGT	CCCTGGATCA	GGCAGGCAAC	CTTACATCAA	ATACAATCAC
		CCGCTAAAAA				
31321	TACAGTCAGC	TCAGGCGCCC	TAACCATGGC	CACAACTTCG	CCTTTGGTGG	TCTCTGACAA
31381	CACTCTTACC	ATGCAATCAC	AAGCACCGCT	AACCGTGCAA	GACTCAAAAC	TTAGCATTGC
31441	TACCAAAGAG	CCACTTACAG	TGTTAGATGG	AAAACTGGCC	CTGCAGACAT	CAGCCCCCCT
		GATAACAACG				
		GCTGTTACCA				
31621	CAAAATTGGC	GGTCCTTTGC	AAGTGGCCAC	CGACTCACAT	GCACTAACAC	TAGGTACTGG
31681	TCAGGGGGTT	GCAGTTCATA	ACAATTTGCT	ACATACAAAA	GTTACAGGCG	CAATAGGGTT
31741	TGATACATCT	GGCAACATGG	AACTTAAAAC	TGGAGATGGC	CTCTATGTGG	ATAGCGCCGG
		AAACTACATA				
31861	CGCAATAACA	ATTAACGCTG	GAAAAGGGTT	GGAATTTGAA	ACAGACTCCT	CAAACGGAAA
31921	TCCCATAAAA	ACAAAAATTG	GATCAGGCAT	ACAATATAAT	ACCAATGGAG	CTATGGTTGC
31981	AAAACTTGGA	ACAGGCCTCA	GTTTTGACAG	CTCCGGAGCC	ATAACAATGG	GCAGCATAAA
32041	CAATGACAGA	CTTACTCTTT	GGACAACACC	AGACCCATCC	CCAAATTGCA	GAATTGCTTC
32101	AGATAAAGAC	TGCAAGCTAA	CTCTGGCGCT	AACAAAATGT	GGCAGTCAAA	TTTTGGGCAC
32161	TGTTTCAGCT	TTGGCAGTAT	CAGGTAATAT	GGCCTCCATC	AATGGAACTC	TAAGCAGTGT
32221	AAACTTGGTT	CTTAGATTTG	ATGACAACGG	AGTGCTTATG	TCAAATTCAT	CACTGGACAA
32281	ACAGTATTGG	AACTTTAGAA	ACGGGGACTC	CACTAACGGT	CAACCATACA	CTTATGCTGT
32341	TGGGTTTATG	CCAAACCTAA	AAGCTTACCC	AAAAACTCAA	AGTAAAACTG	CAAAAAGTAA
32401	TATTGTTAGO	CAGGTGTATC	TTAATGGTGA	CAAGTCTAAA	CCATTGCATT	TTACTATTAC
32461	GCTAAATGGA	ACAGATGAAA	CCAACCAAGT	AAGCAAATAC	TCAATATCAT	TCAGTTGGTC
32521	CTGGAACAGT	GGACAATACA	CTAATGACAA	ATTTGCCACC	AATTCCTATA	CCTTCTCCTA
32581	CATTGCCCAG	GAATAAAGAA	TCGTGAACCT	GTTGCATGTT	ATGTTTCAAC	GTGTTTATTT
32641	TTCAATTGCA	GAAAATTTCA	AGTCATTTT	CATTCAGTAG	TATAGCCCCA	CCACCACATA
32701	GCTTATACTA	ATCACCGTAC	CTTAATCAAA	CTCACAGAAC	CCTAGTATTC	AACCTGCCAC
,_,						

FIG. 7M

	CTCCCTCCCA					
	TATCATGGGT					
	AACGCTCATC					
	CCAGCTGCTG					
	AAGTCCACGC					
	GCAGCAGCGC					
	CAGTGGTCTC					
	CACAGCAGCG					
	TATTGTTTAA					
	AACCCACGTG					
33361	CGCTGGACAT	AAACATTACC	TCTTTTGGCA	TGTTGTAATT	CACCACCTCC	CGGTACCATA
33421	TAAACCTCTG	ATTAAACATG	GCGCCATCCA	CCACCATCCT	AAACCAGCTG	GCCAAAACCT
33481	GCCCGCCGGC	TATGCACTGC	AGGGAACCGG	GACTGGAACA	ATGACAGTGG	AGAGCCCAGG
33541	ACTCGTAACC	ATGGATCATC	ATGCTCGTCA	TGATATCAAT	GTTGGCACAA	CACAGGCACA
33601	CGTGCATACA	CTTCCTCAGG	ATTACAAGCT	CCTCCCGCGT	CAGAACCATA	TCCCAGGGAA
33661	CAACCCATTC	${\tt CTGAATCAGC}$	GTAAATCCCA	CACTGCAGGG	AAGACCTCGC	ACGTAACTCA
33721	CGTTGTGCAT	TGTCAAAGTG	${\tt TTACATTCGG}$	GCAGCAGCGG	ATGATCCTCC	AGTATGGTAG
33781	CGCGTGTCTC	TGTCTCAAAA	GGAGGTAGGC	GATCCCTACT	GTACGGAGTG	CGCCGAGACA
33841	ACCGAGATCG	TGTTGGTCGT	AGTGTCATGC	${\tt CAAATGGAAC}$	GCCGGACGTA	GTCATATTTC
33901	CTGAAGCAAA	ACCAGGTGCG	GGCGTGACAA	ACAGATCTGC	GTCTCCGGTC	TCGTCGCTTA
33961	GCTCGCTCTG	TGTAGTAGTT	GTAGTATATC	CACTCTCTCA	AAGCATCCAG	GCGCCCCTG
34021	GCTTCGGGTT	CTATGTAAAC	TCCTTCATGC	GCCGCTGCCC	TGATAACATC	CACCACCGCA
34081	GAATAAGCCA	CACCCAGCCA	ACCTACACAT	TCGTTCTGCG	AGTCACACAC	GGGAGGAGCG
34141	GGAAGAGCTG	GAAGAACCAT	GTTTTTTTT	TTTATTCCAA	AAGATTATCC	AAAACCTCAA
34201	AATGAAGATC	TATTAAGTGA	ACGCGCTCCC	CTCCGGTGGC	GTGGTCAAAC	TCTACAGCCA
34261	AAGAACAGAT	AATGGCATTT	GTAAGATGTT	GCACAATGGC	TTCCAAAAGG	CAAACTGCCC
34321	TCACGTCCAA	GTGGACGTAA	AGGCTAAACC	CTTCAGGGTG	AATCTCCTCT	ATAAACATTC
34381	CAGCACCTTC	AACCATGCCC	AAATAATTTT	CATCTCGCCA	CCTTATCAAT	ATGTCTCTAA
34441	GCAAATCCCG	AATATTAAGT	CCGGCCATTG	TAAAAATCTG	CTCCAGAGCG	CCCTCCACCT
34501	TCAGCCTCAA	GCAGCGAATC	ATGATTGCAA	AAATTCAGGT	TCCTCACAGA	CCTGTATAAG
34561	ATTCAAAAGC	GGAACATTAA	CAAAAATACC	GCGATCCCGT	AGGTCCCTTC	GCAGGGCCAG
34621	CTGAACATAA	TCGTGCAGGT	CTGCACGGAC	CAGCGCGGCC	ACTTCCCCGC	CAGGAACCAT
34681	GACAAAAGAA	CCCACACTGA	TTATGACACG	CATACTCGGA	GCTATGCTAA	CCAGCGTAGC
34741	CCCGATGTAA	GCTTGTTGCA	TGGGCGGCGA	TATAAAATGC	AAGGTACTGC	TCAAAAAATC
34801	AGGCAAAGCC	TCGCGCAAAA	AAGCAAGCAC	ATCGTAGTCA	TGCTCATGCA	GATAAAGGCA
34861	GGTAAGTTCC	GGAACCACCA	CAGAAAAAGA	CACCATTTTT	CTCTCAAACA	TGTCTGCGGG
34921	TTCCTGCATA	AACACAAAAT	AAAATAACAA	АААААААА	ACATTTAAAC	ATTAGAAGCC
34981	TGTNTTACAA	CAGGAAAAAC	AACCCTTATA	AGCATAAGAC	GGACTACGGC	CATGCCGGCG
35041	TGACCGTAAA	AAAACTGGTC	ACCGTGATTA	AAAAGCACCA	CCGACAGTTC	CTCGGTCATG
35101	TCCGGAGTCA	TAATGTAAGA	CTCGGTAAAC	ACATCAGGTT	GGTTAACATC	GGTCAGTGCT
35161	AAAAAGCGAC	CGAAATAGCC	CGGGGGAATA	CATACCCGCA	GGCGTAGAGA	CAACATTACA
35221	GCCCCCATAG	GAGGTATAAC	ATAATTAAAA	GGAGAGAAAA	ACACATAAAC	ACCTGAAAAA

FIG. 7N

35281	CCCTCCTGCC	TAGGCAAAAT	AGCACCCTCC	CGCTCCAGAA	CAACATACAG	CGCTTCCACA
		TAACAGTCAG				
		CTCAATCAGT				
		ATGACGTAAC				
		AGAAACGAAA				
		ACGTCACTTC				
		TAAAACCTAC				
		TAAAACCTAC				
25701	רשריר <i>א</i> ררירירי	TUATTATUAT	ATTGGCTTCA	WICCUMMIN	1300 *******	

				GAAGCCAATA		
				GCGGGTGACG		
				GCGACGGATG		
				TTTTCGCGCG		
				CCATTTTCGC		
301	AGTGAAATCT	GAATAATTTT	GTGTTACTCA	TAGCGCGTAA	TATTTGTCTA	GGGCCGCGGG
361	GACTTTGACC	GTTTACGTGG	AGACTCGCCC	AGGTGTTTTT	CTCAGGTGTT	TTCCGCGTTC
421	CGGGTCAAAG	TTGGCGTTTT	ATTATTATAG	TCAGCTGACG	TGTAGTGTAT	TTATACCCGG
481	TGAGTTCCTC	AAGAGGCCAC	TCTTGAGTGC	CAGCGAGTAG	AGTTTTCTCC	TCCGAGCCGC
541	TCCGACACCG	GGACTGAAAA	TGAGACATAT	TATCTGCCAC	GGAGGTGTTA	TTACCGAAGA
601	AATGGCCGCC	AGTCTTTTGG	ACCAGCTGAT	CGAAGAGGTA	CTGGCTGATA	ATCTTCCACC
661	TCCTAGCCAT	TTTGAACCAC	CTACCCTTCA	CGAACTGTAT	GATTTAGACG	TGACGGCCCC
				GATTTTTCCC		
781	GCAGGAAGGG	ATTGACTTAC	TCACTTTTCC	GCCGGCGCCC	GGTTCTCCGG	AGCCGCCTCA
841	CCTTTCCCGG	CAGCCCGAGC	AGCCGGAGCA	GAGAGCCTTG	GGTCCGGTTT	CTATGCCAAA
901	CCTTGTACCG	GAGGTGATCG	ATCTTACCTG	CCACGAGGCT	GGCTTTCCAC	CCAGTGACGA
961	CGAGGATGAA	GAGGGTGAGG	AGTTTGTGTT	AGATTATGTG	GAGCACCCCG	GGCACGGTTG
				GGGGGACCCA		
				CAGTAAGTGA		
1141	TAGAGTGGTG	GGTTTGGTGT	GGTAATTTTT	TTTTTAATTT	TTACAGTTTT	GTGGTTTAAA
1201	GAATTTTGTA	TTGTGATTTT	TTTAAAAGGT	CCTGTGTCTG	AACCTGAGCC	TGAGCCCGAG
1261	CCAGAACCGG	AGCCTGCAAG	ACCTACCCGC	CGTCCTAAAA	TGGCGCCTGC	TATCCTGAGA
1321	CGCCCGACAT	CACCTGTGTC	TAGAGAATGC	AATAGTAGTA	CGGATAGCTG	TGACTCCGGT
				GTGGTCCCGC		
				GTGGAATGTA		
				CGCCCCAGGC		
				GAATGAGTTG		
				AATGGGGCGG		
				CTCATGGAGG		
				AGCTCTAACA		
1801				GTCTGCAGAA		
1861				GAGCTGTTTG		
				ACTTTGGATT		
				AAGGATAAAT		
				ATGCATCTGT		
				CGCCCGGCGA		
				CGGCAGGAGC		
				TACAGGTGGC		
				AGGGGCTAAA		
				ATCTAGCTTT		
				AGGATAATTG		
				CCACTTACTG		
				TGGCACTTAG		
				GCTACATTTC		
				GATGTAGCAT		
				ATGTAAGGTT		
2761	GTACGGTTTT	CCTGGCCAAT	ACCAACCTTA	TCCTACACGG	TGTAAGCTTC	TATGGGTTTA
2821	ACAATACCTG	TGTGGAAGCC	TGGACCGATG	TAAGGGTTCG	GGGCTGTGCC	TTTTACTGCT
2881	GCTGGAAGGG	GGTGGTGTGT	CGCCCCAAAA	GCAGGGCTTC	AATTAAGAAA	TGCCTCTTTG
2941	AAAGGTGTAC	CTTGGGTATC	CTGTCTGAGG	GTAACTCCAG	GGTGCGCCAC	AATGTGGCCT
3001	CCGACTGTGG	TTGCTTCATG	CTAGTGAAAA	GCGTGGCTGT	GATTAAGCAT	AACATGGTAT
3061	GTGGCAACTG	CGAGGACAGG	GCCTCTCAGA	TGCTGACCTG	CTCGGACGGC	AACTGTCACC
3121	TGCTGAAGAC	CATTCACGTA	GCCAGCCACT	CTCGCAAGGC	CTGGCCAGTG	TTTGAGCATA
3181	ACATACTGAC	CCGCTGTTCC	TTGCATTTGG	GTAACAGGAG	GGGGGTGTTC	CTACCTTACC
3241	AATGCAATTT	GAGTCACACT	AAGATATTGC	TTGAGCCCGA	GAGCATGTCC	AAGGTGAACC

		000000000000000000000000000000000000000	3.CC3.EC3.3C3	TCTGGAAGGT	CCTGAGGTAC	GATGAGACCC
3301	TGAACGGGGT	GTTTGACATG	ACCATGAAGA	GTAAACATAT	TAGGAACCAG	CCTGTGATGC
3361	GCACCAGGTG	CAGACCCTGC	ACCCCCCAMC	ACTTGGTGCT	GCCCTGCACC	CGCGCTGAGT
3421	TGGATGTGAC	CGAGGAGCTG	AGGCCCGATC	CUNCUCNNAT	CTCTCCCCCT	GGCTTAAGGG
3481	TTGGCTCTAG	CGATGAAGAT	ACAGATTGAG	GTACTGAAAT	ጥልጥርጥርጥጥጥ	GCAGCAGCCG
3541	TGGGAAAGAA	TATATAAGGT	GGGGGTCTTA	TGTAGTTTTG	CACCTCATAT	TTGACAACGC
3601	CCGCCGCCAT	GAGCACCAAC	TCGTTTGATG	GAAGCATTGT	CMCCACCATAT	CATCCTCCCC
3661	GCATGCCCCC	ATGGGCCGGG	GTGCGTCAGA	ATGTGATGGG	CICCAGCAII	ACCCCCTTGG
3721	CCGTCCTGCC	CGCAAACTCT	ACTACCTTGA	CCTACGAGAC	CGIGICIGGA	ATTCTCACTC
3781	AGACTGCAGC	CTCCGCCGCC	GCTTCAGCCG	CTGCAGCCAC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCCCATG
3841	ACTTTGCTTT	CCTGAGCCCG	CTTGCAAGCA	GTGCAGCTTC	CCGTTCATCC	A A TICTIC CTTTT
3901	ACAAGTTGAC	GGCTCTTTTG	GCACAATTGG	ATTCTTTGAC	CCGGGAACTI	MAIGICGIII
3961	CTCAGCAGCT	GTTGGATCTG	CGCCAGCAGG	TTTCTGCCCT	GAAGGCTTCC	AACCAACTCT
4021	ATGCGGTTTA	AAACATAAAT	AAAAAACCAG	ACTCTGTTTG	GATTTGGATC	AAGCAAGIGI
4081	CTTGCTGTCT	TTATTTAGGG	GTTTTGCGCG	CGCGGTAGGC	CCGGGACCAG	CGGICICGGI
4141	CGTTGAGGGT	CCTGTGTATT	TTTTCCAGGA	CGTGGTAAAG	GTGACTCTGG	MCA MCCMCCC
4201	ACATGGGCAT	AAGCCCGTCT	CTGGGGTGGA	GGTAGCACCA	CTGCAGAGCT	TCATGCTGCG
4261	GGGTGGTGTT	GTAGATGATC	CAGTCGTAGC	AGGAGCGCTG	GGCGTGGTGC	CTAAAAATGT
4221	CTTTCACTAC	CAAGCTGATT	GCCAGGGGCA	GGCCCTTGGT	GTAAGTGTTT	ACAAAGCGGI
4201	TA ACCTICICA	TCCCTCCATA	CGTGGGGATA	TGAGATGCAT	CTTGGACTGT	ATTTTIAGGI
4441	ጥርርርጥልጥርጥጥ	CCCAGCCATA	TCCCTCCGGG	GATTCATGTT	GTGCAGAACC	ACCAGCACAG
4501	ጥርሞልጥርርርርጥ	CCACTTCCCA	AATTTGTCAT	GTAGCTTAGA	AGGAAATGUG	TGGAAGAACI
1561	TOCAGACGCC	CTTGTGACCT	CCAAGATTTT	CCATGCATTC	GTCCATAATG	ATGGCAATGG
4621	CCCCACGGGC	CCCCCCCTGG	GCGAAGATAT	TTCTGGGATC	ACTAACGTCA	TAGTIGIGIT
4601	CCACCATGAG	ATCCTCATAG	GCCATTTTTA	CAAAGCGCGG	GCGGAGGGTG	CCAGACTGCG
4711	ርጥለጥለ አጥርርጥ	ጥርርልጥርርርር	CCAGGGGCGT	AGTTACCCTC	ACAGATTTGC	ATTICCCACG
4001	ርጥጥጥር እርጥጥር	ACATGGGGGG	ATCATGTCTA	CCTGCGGGGC	GATGAAGAAA	ACGGTTTCCG
1061	CCCTACCCCA	CATCACCTGG	GAAGAAAGCA	GGTTCCTGAG	CAGCTGCGAC	TTACCGCAGC
1921	CCCTCCCCCCC	GTAAATCACA	CCTATTACCG	GGTGCAACTG	GTAGTTAAGA	GAGCIGCAGC
4091	ጥር ርርርር ርጥር ልጥር	CCTGAGCAGG	GGGGCCACTT	CGTTAAGCAT	GTCCCTGACT	CGCATGTTT
E0/11	CCCTCACCAA	ATCCGCCAGA	AGGCGCTCGC	CGCCCAGCGA	TAGCAGTTCT	TGCAAGGAAG
E101	C y y y C. փորդուրդ	CAACGGTTTG	AGACCGTCCG	CCGTAGGCAT	GCTTTTGAGC	GTTTGACCAA
E161	CCACTTCCAC	CCCCTCCCAC	AGCTCGGTCA	CCTGCTCTAC	GGCATCTCGA	TCCAGCATAT
E221	CTCCTCCTTT	CCCCCCTTCC	GGCGGCTTTC	GCTGTACGGC	AGTAGTCGGT	GCTCGTCCAG
E201	ACCCCCCACC	CTCATCTCTT	TCCACGGGCG	CAGGGTCCTC	GTCAGCGTAG	1C1GGG1CAC
E2/1	CCTCAACCCC	TECECTCEGG	GCTGCGCGCT	GGCCAGGGTG	CGCTTGAGGC	TGGTCCTGCT
5401	CCTCCTCAAC	CGCTGCCGGT	CTTCGCCCTG	CGCGTCGGCC	AGGTAGCATT	TGACCAIGGI
5/61	ርጥሮ እጥ እርጥሮር	AGCCCCTCCG	CGGCGTGGCC	CTTGGCGCGC	AGCTTGCCCT	TGGAGGAGGC
5401	CCCCCACGAG	CCCCACTCCA	GACTTTTGAG	GGCGTAGAGC	TTGGGCGCGA	GAAATACCGA
5521	TTCCCCCCAC	TAGGCATCCG	CGCCGCAGGC	CCCGCAGACG	GTCTCGCATT	CCACGAGCCA
2201	CCTCACCTCT	CCCCCTTCGG	GGTCAAAAAC	CAGGTTTCCC	CCATGCTTTT	TGATGCGTTT
2041	CETTACCTCT	GTTTCCATGA	GCCGGTGTCC	ACGCTCGGTG	ACGAAAAGGC	TGTCCGTGTC
5/01	CITACCICIO	CACTTCACAC	GCCTGTCCTC	GAGCGGTGTT	CCGCGGTCCT	CCTCGTATAG
2/61	CCCGTATACA	CACTIGAGAG	CAAAGGCTCG	CGTCCAGGCC	AGCACGAAGG	AGGCTAAGTG
5821	AAACTCGGAC	CACICIGAGA	CCACTACCC	CTCCACTCGC	TCCAGGGTGT	GAAGACACAT
5881	GGAGGGGTAG	CGGICGIIGI	CCACIAGGGG	TCCTTTCTAG	GTGTAGGCCA	CGTGACCGGG
5941	GTCGCCCTCT	CCCCCCCC	A A A A CCCCCCT	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TCGTCCTCAC	TCTCTTCCGC
6001	TGTTCCTGAA	GGGGGGCIAI	AAAAGGGGG1	. полочение. Полочение	СТСТСАВАВА	CGGGCATGAC
6061	ATCGCTGTCT	CCGAGGGCCA	CTGIIGGG	CCACCACCAT	TTCATATTCA	CCTGGCCCGC
6121	TTCTGCGCTA	AGATTGTCAG	TITICCAAAAA	CGWGGWGGW	A A CA CA A T CT	TTTTGTTGTC
6181	GGTGATGCCT	TIGAGGGIGG	CCGCATCCAT	CIGGICAGAA	2 A A C T T C C C C A	TTTTGTTGTC TGGAGCGCAG
6241	AAGCTTGGTG	GCAAACGACC	CGTAGAGGGG	, GIIGGACAGC	ATCTTGGCC	CCACGTATTC
6301	GGTTTGGTTI	TTGTCGCGAT		, CTIGGCCGC	, WIGILIYACI	GCACGTATTC
6361	GCGCGCAACG	CACCGCCATT	CGGGAAAGAC	, GG1GG1GCGC	, 10010000C	CCAGGTGCAC
6421	. GCGCCAACCG	CGGTTGTGCA	GGGTGACAAC	OTCAACGCTG	GIGGCIACCI	CTCCGCGTAG
6481	GCGCTCGTTG	GTCCAGCAGA	GGCGGCCGCC	CITGCGCGAG	, CAGAAIGGCC	GTAGGGGGTC
6541	TAGCTGCGTC	: TCGTCCGGGG	GGTCTGCGTC	CACGGTAAAG	, ACCCCGGGLA	GCAGGCGCGC

6601	GTCGAAGTAG	TCTATCTTGC	ATCCTTGCAA	GTCTAGCGCC	TGCTGCCATG	CGCGGGCGGC
6661	AAGCGCGCGC	TCGTATGGGT	TGAGTGGGGG	ACCCCATGGC	ATGGGGTGGG	TGAGCGCGGA
				GAGGGGCTCT		
6781	AGGGTAGCAT	CTTCCACCGC	GGATGCTGGC	GCGCACGTAA	TCGTATAGTT	CGTGCGAGGG
6841	AGCGAGGAGG	TCGGGACCGA	GGTTGCTACG	GGCGGGCTGC	TCTGCTCGGA	AGACTATCTG
				GGTTGGACGC		
6961	GTCTGTGAGA	CCTACCGCGT	CACGCACGAA	GGAGGCGTAG	GAGTCGCGCA	GCTTGTTGAC
				GCAGTAGTCC		
7081	ATACTTATCC	TGTCCCTTTT	TTTTCCACAG	CTCGCGGTTG	AGGACAAACT	CTTCGCGGTC
7141	TTTCCAGTAC	TCTTGGATCG	GAAACCCGTC	GGCCTCCGAA	CGGTAAGAGC	CTAGCATGTA
7201	GAACTGGTTG	ACGGCCTGGT	AGGCGCAGCA	TCCCTTTTCT	ACGGGTAGCG	CGTATGCCTG
7261	CGCGGCCTTC	CGGAGCGAGG	TGTGGGTGAG	CGCAAAGGTG	TCCCTGACCA	TGACTTTGAG
7321	GTACTGGTAT	TTGAAGTCAG	TGTCGTCGCA	TCCGCCCTGC	TCCCAGAGCA	AAAAGTCCGT
				GAAGGTGACA		
7441	CGCGCGAGGC	ATAAAGTTGC	GTGTGATGCG	GAAGGGTCCC	GGCACCTCGG	AACGGTTGTT
7501	AATTACCTGG	GCGGCGAGCA	CGATCTCGTC	AAAGCCGTTG	ATGTTGTGGC	CCACAATGTA
7561	AAGTTCCAAG	AAGCGCGGGA	TGCCCTTGAT	GGAAGGCAAT	TTTTTAAGTT	CCTCGTAGGT
7621	GAGCTCTTCA	GGGGAGCTGA	GCCCGTGCTC	TGAAAGGGCC	CAGTCTGCAA	GATGAGGGTT
7681	GGAAGCGACG	AATGAGCTCC	ACAGGTCACG	GGCCATTAGC	ATTTGCAGGT	GGTCGCGAAA
7741	GGTCCTAAAC	TGGCGACCTA	TGGCCATTTT	TTCTGGGGTG	ATGCAGTAGA	AGGTAAGCGG
				CGCGGCTAGG		
				CATGAAGGGC		
				GGTGACAAAG		
				CCACCAATTG		
8041	GTGAAAGTAG	AAGTCCCTGC	GACGGGCCGA	ACACTCGTGC	TGGCTTTTGT	AAAAACGTGC
				ATCCTGCACG		
				CTCGCCTGGC		
8221	TACTTCGGCT	GCTTGTCCTT	GACCGTCTGG	CTGCTCGAGG	GGAGTTACGG	TGGATCGGAC
8281	CACCACGCCG	CGCGAGCCCA	AAGTCCAGAT	GTCCGCGCGC	GGCGGTCGGA	GCTTGATGAC
8341	AACATCGCGC	AGATGGGAGC	TGTCCATGGT	CTGGAGCTCC	CGCGGCGTCA	GGTCAGGCGG
				GGTCAGGGCG		
8461	ССТААТТТСС	AGGGGCTGGT	TGGTGGCGGC	GTCGATGGCT	TGCAAGAGGC	CGCATCCCCG
8521	CCCCCCCCACT	ACGGTACCGC	GCGGCGGCG	GTGGGCCGCG	GGGGTGTCCT	TGGATGATGC
8581	ATCTABAACC	CCTCACCCCG	GCGAGCCCCC	GGAGGTAGGG	GGGGCTCCGG	ACCCGCCGGG
8641	ACACCCCCCA	CCCCCACCTC	GGCGCCGCGC	GCGGGCAGGA	GCTGGTGCTG	CGCGCGTAGG
8701	TTCCTCCCGA	ACGCGACGAC	GCGGCGGTTG	ATCTCCTGAA	TCTGGCGCCT	CTGCGTGAAG
				GAGAGTTCGA		
				ACGTCTCCTG		
0021	TCCCCCATCA	ACTCCTCCAT	CTCTTCCTCC	TGGAGATCTC	CGCGTCCGGC	TCGCTCCACG
90/1	CTCCCCCCCA	CCTCCTTCCA	AATCCGGGCC	ATGAGCTGCG	AGAAGGCGTT	GAGGCCTCCC
9001				CCTTCGGCAT		
9061				AAGACGGCGT		
				GCCACGAAGA		
				TCAAGGCGCT		
				GCCGACACGG		
				TCGCGCTCAA		
930I	CGGATGAGCT	CGGCGACAGT	CAMARCOCCC	TCCCCTTCTT	COULTACAGG	CCCCCCTCCC
320T	TUTTUTTUAA	CACCCCCCCC	ACCACCCCCC	ACCGGGAGGC	CITCITCIGG	CCCCACCATAGG
9481	ATUTUCCUGC	CCCCCCCCCC	CATGGTCTCG	GTGACGGCGC	CCCCCCCCCC	CCCATCCCCC
324T	AGTTGGAAGA	CGCCGCCCGT	CATGICCCGG	TTATGGGTTG	TA COTA CTCC	GCCWIGCGGC
A001	AGGGATACGG	CGCTAACGAT	GUATUTUAAU	AATTGTTGTG	CCACAAACCC	COCCOCCOAGG
				GAAAACCTCT		
				GCGGGCGCA		
9781	TTTCTGGCGG	AGGTGCTGCT	GATGATGTAA	TTAAAGTAGG	magages ages	ACGGCGGATG
9841	GTCGACAGAA	GCACCATGTC	CTTGGGTCCG	GCCTGCTGAA	TGCGCAGGCG	GTCGGCCATG

PCT/US02/32512

60/92

WO 03/031588

0001	CCCCACCCTT	CGTTTTGACA	TCGGCGCAGG	TCTTTGTAGT	AGTCTTGCAT	GAGCCTTTCT
9901	ACCCCCA COT	CTTCTTCTCC	T	CCTGCATCTC	TTGCATCTAT	CGCTGCGGCG
10021	ACCGCCACTI	TTGGCCGTAG	CTGCCGCCCCT	CTTCCTCCCA	TGCGTGTGAC	CCCGAAGCCC
10021	GCGGCGGAGI	GAAGCAGGGC	TAGGTCGGCG	ACAACGCGCT	CGGCTAATAT	GGCCTGCTGC
10081	CTCATCGGCI	GGGTAGACTG	CAACTCATCC	ATGTCCACAA	AGCGGTGGTA	TGCGCCCGTG
10141	ACCTGCGTGA	AAGTGCAGTT	CCCCATAACC	CACCAGTTAA	CGGTCTGGTG	ACCCGGCTGC
10201	TIGATGGTGT	TGTACCTGAG	ACCCCACTAA	CCCCTCGAGT	CAAATACGTA	GTCGTTGCAA
10261	GAGAGCTCGG	GGTACTGGTA	TCCCACCAAA	AACTGCGGCG	GCGGCTGGCG	GTAGAGGGGC
10321	GTCCGCACCA	TGGCCGGGGC	MCCCCCCCCCCC	ACATCTTCCA	ACATAAGGCG	ATGATATCCG
10381	CAGCGTAGGG	TGGCCGGGGC	CCTCATCCC	CCCCCCCCTGG	TGGAGGCGCG	CGGAAAGTCG
10441	TAGATGTACC	TCCAGATGTT	CCCCACCCCC	A A A A A CTCCT	CCATGGTCGG	GACGCTCTGG
10501	CGGACGCGGT	GCGCGCAATC	CERCACCCEC	TACACCCTCC	AAAAGGAGAG	CCTGTAAGCG
10561	CCGGTCAGGC	CGTGGTCTGG	MCCAMA A AMT	CCCAACCCTA	TCATGGCGGA	CGACCGGGGT
10621	GGCACTCTTC	TATCCGGCCG	TGGATAAATT	TOCATOCCET	TACCGCCCGC	GTGTCGAACC
10681	TCGAGCCCCG	ACGTCAGACA	ACCCCCCACO	CCTCCTTTTTC	CCTTCCTTCC	AGGCGCGGCG
10741	CAGGTGTGCG	TAGCTTTTTT	ACGGGGGAG1	CCCCCCCACC	CTA ACCGGTT	AGGCTGGAAA
10801	GCTGCTGCGC	TAGCTTTTTT	GGCCAC TGGC	CCCCCACCCT	TATTTTCCAA	GGGTTGAGTC
10861	GCGAAAGCAT	TAAGTGGCTC	GCTCCCTGTA	GCCGGAGGG1	CCCAACGGGG	CTTTCCCTCC
10921	GCGGGACCCC	CGGTTCGAGT	CICGGACCGG	CCGGAC1GCG	CCCACCACCC	CCTTTTTTTTCC
10981	CCGTCATGCA	AGACCCCGCT	TGCAAATTCC	AMOCCCCCCC	CTCCTCAGCA	GCGGCAAGAG
11041	TTTTCCCAGA	TGCATCCGGT	GCTGCGGCAG	MACCOMMONTO	CTCCTCAGCA	ACCACCICIO
11101	CAAGAGCAGC	GGCAGACATG	CAGGGCACCC	CAMMACCAAC	CCCCCCCCCCC	CCGGGCCCGG
11161	ACATCCGCGG	TTGACGCGGC	AGCAGATGGT	GATTACGAAC	TACCACCECC	CTCTCCTGAG
11221	CACTACCTGG	ACTTGGAGGA	GGGCGAGGGC	TGGCGCGC	CCMACCMCCC	CCCCCAGAAC
11281	CGGTACCCAA	GGGTGCAGCT	GAAGCGTGAT	ACGCGTGAGG	CCCAMCCAAA	CTTCCACCCA
11341	CTGTTTCGCG	ACCGCGAGGG	AGAGGAGCCC	GAGGAGATGC	MOCCOCO CO	CCACTTTCAC
11401	GGGCGCGAGC	TGCGGCATGG	CCTGAATCGC	GAGCGGTTGC	TGCGCGAGGA	CCACCTCCTA
11461	CCCGACGCGC	GAACCGGGAT	TAGTCCCGCG	CGCGCACACG	TGGCGGCCGC	CGACCIGGIA
11521	ACCGCATACG	AGCAGACGGT	GAACCAGGAG	ATTAACTTTC	AAAAAAGCTT	TAACAACCAC
11501	CTCCCTACCC	TTGTGGCGCG	CGAGGAGGTG	GCTATAGGAC	TGATGUATUT	GIGGGACIII
11611	CTAACCCCCCC	TCCACCAAAA	CCCAAATAGC	AAGCCGCTCA	TGGCGCAGCT	GITCCITATA
11701	CTCCACCACA	CCACCCACAA	CGAGGCATTC	AGGGATGCGC	TGCTAAACAT	AGTAGAGCCC
11761	CACCCCCCCT	CCCTCCTCGA	TTTGATAAAC	ATCCTGCAGA	GCATAGTGGT	GCAGGAGCGC
11021	አርርምምር አርርር	TCCCTCACAA	GGTGGCCGCC	ATCAACTATT	CCATGCTTAG	CCTGGGCAAG
11881	TTTTACGCCC	GCAAGATATA	CCATACCCCT	TACGTTCCCA	TAGACAAGGA	GGTAAAGATC
11041	CACCCCTTCT	ACATGCGCAT	GGCGCTGAAG	GTGCTTACCT	TGAGCGACGA	CCTGGGCGTT
12001	ጥአጥሮርር ል ልሮር	AGCGCATCCA	CAAGGCCGTG	AGCGTGAGCC	GGCGGCGCGA	GCTCAGCGAC
12061	CCCCACCTCA	TOCACACCOT	GCAAAGGGCC	CTGGCTGGCA	CGGGCAGCGG	CGATAGAGAG
12121	CCCCACTCCT	ACTITICACCC	GGGCGCTGAC	CTGCGCTGGG	CCCCAAGCCG	ACGCGCCCTG
12101	CACCCACCTG	GGGCCGGACC	TGGGCTGGCG	GTGGCACCCG	CGCGCGCTGG	CAACGTCGGC
122/1	CCCCTCCACC	AATATGACGA	GGACGATGAG	TACGAGCCAG	AGGACGGCGA	GTACTAAGCG
12301	CTCATCTTTC	TCATCAGATG	ATGCAAGACG	CAACGGACCC	GGCGGTGCGG	GCGGCGCTGC
12261	ACACCCACCC	CTCCGGCCTT	AACTCCACGG	ACGACTGGCG	CCAGGTCATG	GACCGCATCA
12421	MCMCCCTCAC	TCCCCCCAAT	CCTGACGCGT	TCCGGCAGCA	GCCGCAGGCC	AACCGGCTCT
12/01	CCCC A A ጥጥርጥ	CCAACCGGTG	GTCCCGGCGC	GCGCAAACCC	CACGCACGAG	AAGGTGCTGG
125/1	CCDTCCTDDDD	CGCGCTGGCC	GAAAACAGGG	CCATCCGGCC	CGACGAGGCC	GGCCTGGTCT
12601	ACCACCCCCT	CCTTCAGCGC	GTGGCTCGTT	ACAACAGCGG	CAACGTGCAG	ACCAACCIGG
12661	እድድርርር ምርርጥ	CCCCCATCTC	CGCGAGGCCG	TGGCGCAGCG	TGAGCGCGCG	CAGCAGCAGG
12721	CCAACCTCCC		GCACTAAACG	CCTTCCTGAG	TACACAGCCC	GCCAACGIGC
12701	CCCCCCCACA	CCACCACTAC	ACCAACTTTG	TGAGCGCACT	GCGGCTAATG	GIGACIGAGA
12041	CACCCCAAAC	TCACCTCTAC	CAGTCTGGGC	CAGACTATT	TTTCCAGACC	AGTAGACAAG
12001	CCCTCCACAC	CCTABACCTC	AGCCAGGCTT	' TCAAAAACTI	GCAGGGGCTG	1666666166
12061	CCCCTCCCAC	* ACCCCACCCC	GCGACCGTGT	' CTAGCTTGCT	GACGCCCAAC	TUGUGUUTGI
12021	MCCMCCMCCT	י אאידא כרכררר	-TTCACGGACA	GTGGCAGCGT	GTCCCGGGAC	ACATACCIAG
12001	CMC & CMMCCM	י כארארית בייארי	CGCGAGGCCA	TAGGTCAGGC	: GCATGTGGAC	GAGCATACTT
13081	BCCACTTGCT	MACARCIGIAC	≱GCCRCGCGC	TGGGGCAGGA	GGACACGGGC	AGCCTGGAGG
13141	TCCAGGAGAT	INCANGIGIC	AGCCGCGCGC			

	CAACCCTAAA					
	ACAGCGAGGA					
	GCGACGGGGT					
	TGTATGCCTC					
	CCGTGAACCC					
	GTTTCTACAC					
	TAGACGACAG					
	AGGCAGAGGC					
13681	GCGCTGCGGC	CCCGCGGTCA	GATGCTAGTA	GCCCATTTCC	AAGCTTGATA	GGGTCTCTTA
13741	CCAGCACTCG	CACCACCCGC	CCGCGCCTGC	TGGGCGAGGA	GGAGTACCTA	AACAACTCGC
	TGCTGCAGCC					
	GCCTAGTGGA					
	GCCCGCGCCC					
	ACGATGACTC					
	CGCACCTTCG					
	AAAAACTCAC					
14161	GCGCGCGCG	ATGTATGAGG	AAGGTCCTCC	TCCCTCCTAC	GAGAGTGTGG	TGAGCGCGGC
14221	GCCAGTGGCG	GCGGCGCTGG	GTTCTCCCTT	CGATGCTCCC	CTGGACCCGC	CGTTTGTGCC
14281	TCCGCGGTAC	CTGCGGCCTA	CCGGGGGGAG	AAACAGCATC	CGTTACTCTG	AGTTGGCACC
	CCTATTCGAC					
	GAACTACCAG					
	CCCGGGGGAG					
	CCTGAAAACC					
	GTTTAAGGCG					
	ATACGAGTGG					
	CCTTATGAAC					
14761	GGAAAGCGAC	ATCGGGGTAA	AGTTTGACAC	CCGCAACTTC	AGACTGGGGT	TTGACCCCGT
	CACTGGTCTT					
	GCTGCCAGGA					
	CAAGCGGCAA					
	CATTCCCGCA					
	GGGCGGGGT					
	CGCGGCAGCC					
15181	CACCTTTGCC	ACACGGGCTG	AGGAGAAGCG	CGCTGAGGCC	GAAGCAGCGG	CCGAAGCTGC
15241	CGCCCCGCT	GCGCAACCCG	AGGTCGAGAA	GCCTCAGAAG	AAACCGGTGA	TCAAACCCCT
	GACAGAGGAC					
15361	GTACCGCAGC	TGGTACCTTG	CATACAACTA	CGGCGACCCT	CAGACCGGAA	TCCGCTCATG
15421	GACCCTGCTT	TGCACTCCTG	ACGTAACCTG	CGGCTCGGAG	CAGGTCTACT	GGTCGTTGCC
15481	AGACATGATG	CAAGACCCCG	TGACCTTCCG	CTCCACGCGC	CAGATCAGCA	ACTTTCCGGT
15541	GGTGGGCGCC	GAGCTGTTGC	CCGTGCACTC	CAAGAGCTTC	TACAACGACC	AGGCCGTCTA
15601	CTCCCAACTC	ATCCGCCAGT	TTACCTCTCT	GACCCACGTG	TTCAATCGCT	TTCCCGAGAA
15661	CCAGATTTTG	GCGCGCCCGC	CAGCCCCCAC	CATCACCACC	GTCAGTGAAA	ACGTTCCTGC
15721	TCTCACAGAT	CACGGGACGC	TACCGCTGCG	CAACAGCATC	GGAGGAGTCC	AGCGAGTGAC
15781	CATTACTGAC	GCCAGACGCC	GCACCTGCCC	CTACGTTTAC	AAGGCCCTGG	GCATAGTCTC
	GCCGCGCGTC					
15901	CAATAACACA	GGCTGGGGCC	TGCGCTTCCC	AAGCAAGATG	TTTGGCGGGG	CCAAGAAGCG
	CTCCGACCAA					
	ACGCGGCCGC					
16081	GCGCAACTAC	ACGCCCACGC	CGCCACCAGT	GTCCACAGTG	GACGCGGCCA	TTCAGACCGT
16141	GGTGCGCGGA	GCCCGGCGCT	ATGCTAAAAT	GAAGAGACGG	CGGAGGCGCG	TAGCACGTCG
16201	CCACCGCCGC	CGACCCGGCA	CTGCCGCCCA	ACGCGCGGCG	GCGGCCCTGC	TTAACCGCGC
16261	ACGTCGCACC	GGCCGACGGG	CGGCCATGCG	GGCCGCTCGA	AGGCTGGCCG	CGGGTATTGT
16321	CACTGTGCCC	CCCAGGTCCA	GGCGACGAGC	GGCCGCCGCA	GCAGCCGCGG	CCATTAGTGC
16381	TATGACTCAG	GGTCGCAGGG	GCAACGTGTA	TTGGGTGCGC	GACTCGGTTA	GCGGCCTGCG
16441	CGTGCCCGTG	CGCACCCGCC	CCCCGCGCAA	CTAGATTGCA	AGAAAAAACT	ACTTAGACTC

PCT/US02/32512

	•					
16501	GTACTGTTGT	ATGTATCCAG	CGGCGGCGGC	GCGCAACGAA	GCTATGTCCA	AGCGCAAAA'I'
16561	CANACARCAC	ATCCTCCACC	TCATCGCGCC	GGAGATCTAT	GGCCCCCGA	ACAAGGAAGA
16621	CCACCATTAC	AAGCCCCGAA	AGCTAAAGCG	GGTCAAAAAG	AAAAAGAAAG	ATGATGATGA
1 6 6 0 1	ጥር እ እርጥጥር እር	CACCAGGTGG	AACTGCTGCA	CGCTACCGCG	CCCAGGCGAC	GGGTACAGIG
16711	CARACCTCCA	CCCCTAAAAC	GTGTTTTGCG	ACCCGGCACC	ACCGTAGTCT	TTACGCCCGG
1 6001	THE RECEPTION	ACCCGCACCT	ACAAGCGCGT	GTATGATGAG	GTGTACGGCG	ACGAGGACCI
10061	CCMTCACCAC	CCCAACGAGC	GCCTCGGGGA	GTTTGCCTAC	GGAAAGCGGC	ATAAGGACAI
16021	CCTCCCCTTC	CCGCTGGACG	AGGGCAACCC	AACACCTAGC	CTAAAGCCCCG	TAACACIGCA
1 (0 0 1	CCXCCTCCTC	CCCCCCCTTG	CACCGTCCGA	AGAAAAGCGC	GGCCTAAAGC	GCGAGICIGG
17041	MC A COTTCCC A	CCCACCGTGC	AGCTGATGGT	ACCCAAGCGC	CAGCGACTGG	AAGATGTCTT
17101	CCARARARC	ACCGTGGAAC	CTGGGCTGGA	GCCCGAGGTC	CGCGTGCGGC	CAATCAAGCA
17101	GGAAAAAA	GGACTGGGCG	TGCAGACCGT	GGACGTTCAG	ATACCCACTA	CCAGTAGCAC
1/101	GGTGGCGCCG	ACCGCCACAG	ACCCCATCGA	GACACAAACG	TCCCCGGTTG	CCTCAGCGGT
17221	CAGTATTGCC	GCGGTGCAGG	CCCTCCCTCC	GGCCGCGTCC	AAGACCTCTA	CGGAGGTGCA
17281	GGCGGATGCC	TGGATGTTTC	CCCTTTCACC	CCCCCGGCGC	CCGCGCGGTT	CGAGGAAGTA
17341	AACGGACCCG	AGCGCGCTAC	TCCCCCAATA	TGCCCTACAT	CCTTCCATTG	CGCCTACCCC
17401	CGGCGCCGCC	GGCTACACCT	ACCCCCCCAG	AAGACGAGCA	ACTACCCGAC	GCCGAACCAC
17461	CGGCTATCGT	CGCCGCCGCC	CTCCCCCTCC	CCACCCCCTG	CTGGCCCCGA	TTTCCGTGCG
17521	CACTGGAACC	CGCCGAAGGAG	CCACCACCCA	CCTCCTCCCA	ACAGCGCGCT	ACCACCCCAG
17581	CAGGGTGGCT	AAGCCGGTCT	GCAGGACCC1	TOCACATATC	CCCCTCACCT	GCCGCCTCCG
17641	CATCGTTTAA	CCGGGATTCC	CACCAACAAC	CCACCCTACC	ACCCCCATGG	CCGGCCACGG
17701	TTTCCCGGTG	CCGGGATTCC	GAGGAAGAAI	GCACCGIAGG	CCCCCCTCCC	ACCGTCGCAT
17761	CCTGACGGGC	GGCATGCGTC	GTGCGCACCA	A CITICA INCCCCC	CCCCCCATTC	GCGCCGTGCC
17821	GCGCGGCGGT	ATCCTGCCCC	TCCTTATTCC	ACTGATCGCC	GCGGCGAIIG	CTTCCATCTC
17881	CGGAATTGCA	TCCGTGGCCT	TGCAGGCGCA	GAGACACTGA	TIMAMAMCAA	CUNTUTUTOUS
17941	GAAAAATCAA	AATAAAAAGT	CTGGACTCTC	ACGCTCGCTT	GGTCCTGTAA	CIMITITUE
18001	GAATGGAAGA	CATCAACTTT	GCGTCTCTGG	CCCCGCGACA	CGGCTCGCGC	TCCCTTCATGG
18061	GAAACTGGCA	AGATATCGGC	ACCAGCAATA	TGAGCGGTGG	CGCCTTCAGC	1GGGGC1CGC
18121	TGTGGAGCGG	CATTAAAAAT	TTCGGTTCCA	CCGTTAAGAA	CTATGGCAGC	AAGGCCIGGA
18181	ACAGCAGCAC	AGGCCAGATG	CTGAGGGATA	AGTTGAAAGA	GCAAAATTI'I'C	CAACAAAAGG
102/1	TOOTAGATOG	CCTGGCCTCT	GGCATTAGCG	GGGTGGTGGA	CCTGGCCAAC	CAGGCAGIGC
10201	አአአአጥአአርኔጥ	TAACAGTAAG	CTTGATCCCC	GCCCTCCCGT	AGAGGAGCCT	CCACCGGCCG
10361	TOCACACACT	GTCTCCAGAG	GGGCGTGGCG	AAAAGCGTCC	GCGCCCCGAC	AGGGAAGAAA
10/71	CTCTCCTCAC	CCAAATAGAC	GAGCCTCCCT	CGTACGAGGA	GGCACTAAAG	CAAGGCCTGC
10401	CCACCACCCC	TOCOLATOGOG	CCCATGGCTA	CCGGAGTGCT	GGGCCAGCAC	ACACCCGTAA
105/1	CCCTCCACCT	CCCTCCCCC	GCCGACACCC	AGCAGAAACC	TGTGCTGCCA	GGCCCGACCG
10601	CCCTTCTTCT	AACCCCTCCT	AGCCGCGCGT	CCCTGCGCCG	CGCCGCCAGC	GGICCGCGAI
10661	CCTTCCCCCC	CCTACCCAGT	GGCAACTGGC	AAAGCACACT	GAACAGCATC	GIGGICIGG
10721	CCCTCCAATC	CCTGAAGCGC	CGACGATGCT	TCTGAATAGC	TAACGTGTCG	TAIGIGIGIC
10701	አጥርጥልጥርሮርጥ	CCATGTCGCC	GCCAGAGGAG	CTGCTGAGCC	GCCGCGCGCC	CGCTTTCCAA
10041	CAMCCCCTACC	CCTTCGATGA	TGCCGCAGTG	GTCTTACATG	CACATCTCGG	GUCAGGACGC
10001	CTCCCACTAC	CTGAGCCCCG	GGCTGGTGCA	GTTTGCCCGC	GCCACCGAGA	CGTACTTCAG
10061	CCMC እ አጥአ እ C	አአርጥጥጥልርልል	ACCCCACGGT	GGCGCCTACG	CAUGACGIGA	CCACAGACCG
10001	COLGUETARC	TTGACGCTGC	GGTTCATCCC	TGTGGACCGT	GAGGATACTG	CGTACTCGTA
19021	GICCCAGCGI	DATODOAGUT DATODOAGUT	СТСТСССТСА	TAACCGTGTG	CTGGACATGG	CTTCCACGTA
19081	CAAGGCGCGCGC	CGCGGCGTGC	TCGACAGGGG	CCCTACTTTT	AAGCCCTACT	CTGGCACTGC
19141	CTTTGACATC	CTGGCTCCCA	ACCCTCCCC	AAATCCTTGC	GAATGGGATG	AAGCTGCTAC
19201	CTACAACGCC	CIGGCICCA	AGGGIGCCC	CCATGACAAC	GAAGACGAAG	TAGACGAGCA
19261	TGCTCTTGAA	ATAAACCIAG	ACCUPATUTCC	CCAGGCGCCT	TATTCTGGTA	TAAATATTAC
19321	AGCTGAGCAG	ADDONANCE C	TOTAL TIGG	TCAAACACCT	AAATATGCCG	ATAAAACATT
19381	AAAGGAGGGT	ATTCAAATAG	GIGICGWAGG	CACCAPCE	АСТСАВАТТА	ATCATGCAGC
19441	. TCAACCTGAA	CCTCAAATAG	CMACCCCAAG	CAPACCAVA	ተልቦርርጥጥቦልባ	ATGCAAAACC
19501	TGGGAGAGTC	CTTAAAAAGA	A A CCC A MMCC	1017337CCT	СРУРАТССТВ	AGCTAGAAAG
19561	. CACAAATGAA	AATGGAGGGC	AAGGCATTCT	**************************************	CAAAA GOAA	GTGATAACTT
19621	. TCAAGTGGAA	ATGCAATTTT	TOTOAACTAC	TGMGGCGMCC	. GUAGGUAAIG	GTGATAACTT
19681	. GACTCCTAAA	GTGGTATTGT	ACAGTGAAGA	TGIMGMIMIA	CANTACCCCAC	ACACTCATAT
19741	TTCTTACATO	CCCACTATTA	AGGAAGGTAA	CICACGAGAA	CIMMIGGGCC	AACAATCTAT

10001	GCCCAACAGG	CCMAAMMACA	TTCCTTTTT > C	CCX CX X DODT	እ <i>ጥ</i> ጥር ርጥር ጥእ እ	ጥርጥል ጥጥል ሮ ል ል
	CAGCACGGGT					
	TTTGCAAGAC					
	AACCAGGTAC					
	TATTGAAAAT					
	GATTAATACA					
	AAAAGATGCT			-		
	GGAAATCAAT					
20281	TTTGCCCGAC	AAGCTAAAGT	ACAGTCCTTC	CAACGTAAAA	ATTTCTGATA	ACCCAAACAC
	CTACGACTAC	-	-			
20401	TGGAGCACGC	TGGTCCCTTG	ACTATATGGA	CAACGTCAAC	CCATTTAACC	ACCACCGCAA
20461	TGCTGGCCTG	CGCTACCGCT	CAATGTTGCT	GGGCAATGGT	CGCTATGTGC	CCTTCCACAT
	CCAGGTGCCT					
20581	CTACGAGTGG	AACTTCAGGA	AGGATGTTAA	CATGGTTCTG	CAGAGCTCCC	TAGGAAATGA
	CCTAAGGGTT					
20701	CCCCATGGCC	CACAACACCG	CCTCCACGCT	TGAGGCCATG	CTTAGAAACG	ACACCAACGA
20761	CCAGTCCTTT	AACGACTATC	TCTCCGCCGC	CAACATGCTC	TACCCTATAC	CCGCCAACGC
20821	TACCAACGTG	CCCATATCCA	TCCCCTCCCG	CAACTGGGCG	GCTTTCCGCG	GCTGGGCCTT
20881	CACGCGCCTT	AAGACTAAGG	AAACCCCATC	ACTGGGCTCG	GGCTACGACC	CTTATTACAC
20941	CTACTCTGGC	TCTATACCCT	ACCTAGATGG	AACCTTTTAC	CTCAACCACA	CCTTTAAGAA
21001	GGTGGCCATT	ACCTTTGACT	CTTCTGTCAG	CTGGCCTGGC	AATGACCGCC	TGCTTACCCC
	CAACGAGTTT					
	CATGACCAAA					
21181	CTTCTATATC	CCAGAGAGCT	ACAAGGACCG	CATGTACTCC	TTCTTTAGAA	ACTTCCAGCC
	CATGAGCCGT					
	ACACCAACAC					
	GGCCTACCCT					
	CCAGAAAAAG					
	GTCCATGGGC					
	GCTAGACATG					
	TGAAGTCTTT					
	CCTGCGCACG					
	ACAGCTGCCG					
	TGTGGGCCAT					
	AAGCTCGCCT					
	GCCTTTGCCT					
	GACCAGCGAC					
	ATTGCTTCTT					
	CCCAACTCGG					
	CCCCAAACTC					
	ATGCTCAACA					
	TTCCTGGAGC					
	TCTTTTTGTC					
	AAATGCTTTT					
	GTTTAAAAAT	-				
	CGATACTGGT					
	AAGTTTTCAC					
	ATCTTGAAGT					
	CAGCACTGGA					
	ATCAGATCCG					
	TGCCTTCCCA					
	AAAAGGTGAC					
	TGCTTAAAAG					
	GAAAACTGAT					
23041	ATCTGCACCA	CATTTCGGCC	CCACCGGTTC	TTCACGATCT	TGGCCTTGCT	AGACTGCTCC

	mma. 000000	GCTGCCCGTT	ጥጥርርርጥርርጥር	ACATCCATTT	CAATCACGTG	CTCCTTATTT
23101	TTCAGCGCGC	TTCCGTGTAG	ACACMMA ACC	TCCCCTTCGA	TCTCAGCGCA	GCGGTGCAGC
23161	ATCATAATGC	AGCCCGTGGG	CTCCTCATCC	TTGTAGGTCA	CCTCTGCAAA	CGACTGCAGG
23221	CACAACGCGC	GGAATCGCCC	CICGIGAIGC	ACABACGTCT	TGTTGCTGGT	GAAGGTCAGC
23281	TACGCCTGCA	GGAATCGCCC	CMICAICGIC	CTCTTCCATA	CGGCCGCCAG	AGCTTCCACT
23341	TGCAACCCGC	GTAGTTTGAA	CERCOCCERC	A CATCGTTAT	CCACGTGGTA	CTTGTCCATC
23401	TGGTCAGGCA	CAGCCTCCAT	COCCUTTI	CYCCCYCYCY	CGATCGGCAC	ACTCAGCGGG
23461	AGCGCGCGCG	TAATTTCACT	THE COCKET CO	CACCCCCCCCTCTT	CCTCTTCCTC	TTGCGTCCGC
23521	TTCATCACCG	CCACTGGGTC	TTCCGCTTCG	ACCCCCCCCA	CTCTCCCCTT	ACCTCCTTTG
23581	ATACCACGCG	CCACTGGGTC	CTCTTCATTC	AGCCGCCACCA	TTTGTAGCGC	CACATCTTCT
23641	CCATGCTTGA	TTAGCACCGG	TGGGTTGCTG	CCCCACCA	CCCCCTCCCC	CTTGGGAGAA
23701	CTTTCTTCCT	CGCTGTCCAC	GATTACCTCT	GGIGAIGGCG	CCCCCCACCT	CGATGGCCGC
23761	GGGCGCTTCT	TTTTCTTCTT	GGGCGCAATG	TOTO ATTO ACT	CTTCCTCGTC	CTCGGACTCG
23821	GGGCTGGGTG	TGCGCGGCAC	CAGCGCGTCT	TGTGATGAGT	CCCCCCCC	CGGGGACGGG
23881	ATACGCCGCC	TCATCCGCTT	TTTTGGGGGGC	CCCCCCCCAC	CCCCTCCCC	CTCGGGGGTG
23941	GACGACACGT	CCTCCATGGT	TGGGGGACGT	A MOUNT COUNTY CO.	CCTATACCCA	GAAAAAGATC
24001	GTTTCGCGCT	GCTCCTCTTC	CCGACTGGCC	ATTICCTICT	CTCACTTCCC	CACCACCGCC
24061	ATGGAGTCAG	TCGAGAAGAA	GGACAGCCTA	ACCGCCCCCT	ACCCACCCCC	CCTTCACGAG
24121	TCCACCGATG	CCGCCAACGC	GCCTACCACC	TTCCCCGTCG	AGGCACCCC	CCACCCCTCA
24181	GAGGAAGTGA	TTATCGAGCA	GGACCCAGGT	TTTGTAAGCG	AAGACGACGA	CCAACAACTC
24241	GTACCAACAG	AGGATAAAA	GCAAGACCAG	GACAACGCAG	AGGCAAACGA	CCTCTTCAAC
24301	GGGCGGGGG	ACGAAAGGCA	TGGCGACTAC	CTAGATGTGG	GAGACGACGI	CCATCTCCCC
24361	CATCTGCAGC	GCCAGTGCGC	CATTATCTGC	GACGCGTTGC	MAGAGCGCAG	CCCCCTACCC
24421	CTCGCCATAG	CGGATGTCAG	CCTTGCCTAC	GAACGCCACC	TATTCTCACC	CENCCCCCEN
24481	CCCAAACGCC	AAGAAAACGG	CACATGCGAG	CCCAACCCGC	GCCTCAACTT	CAACATACCC
24541	TTTGCCGTGC	CAGAGGTGCT	TGCCACCTAT	CACATCTTT	TCCAAAACTG	CAAGATACCC
24601	ርጥን ጥር ርጥር ርር	CTCCCAACCG	CAGCCGAGCG	GACAAGCAGC	1666611666	GCVGGGCGC1
24661	GTCATACCTG	ATATCGCCTC	GCTCAACGAA	GTGCCAAAAA	TCTTTGAGGG	TCTTGGACGC
24721	GACGAGAAGC	GCGCGGCAAA	CGCTCTGCAA	CAGGAAAACA	GCGAAAATGA	AAGTCACTCT
24701	CCXCTCTTCC	ጥርር እ እርጥሮር እ	GCCTGACAAC	GCGCGCCTAG	CCGTACTAAA	ACGCAGCAIC
24041	CACCECACCC	አርጥጥጥርርርርጥ እ	CCCGGCACTT	AACCTACCCC	CCAAGGTCAT	GAGCACAGIC
24001	አጥር እርጥር እር ር	TGATCGTGCG	CCGTGCGCAG	CCCCTGGAGA	GGGATGCAAA	TTTGCAAGAA
24061	CANACACACC	ACCCCCTACC	CGCAGTTGGC	GACGAGCAGC	TAGCGCGCTG	GCTTCAAACG
25421	CCCCACCCTC	CCGACTTGGA	GGAGCGACGC	AAACTAATGA	TGGCCGCAGT	GCTCGTTACC
25001	CTCCACCTTC	ACTGCATGCA	GCGGTTCTTT	GCTGACCCGG	AGATGCAGCG	CAAGCTAGAG
25141	CAAACATTCC	ACTACACCTT	TCGACAGGGC	TACGTACGCC	AGGCCTGCAA	GATCICCAAC
25201	CTCCACCTCT	CCAACCTGGT	CTCCTACCTT	GGAATTTTGC	ACGAAAACCG	CCTTGGGCAA
25261	አ አ ሮርጥርር ጥጥር	ልጥጥርር ACGCT	CAAGGGCGAG	GCGCGCCGCG	ACTACGTCCG	CGACTGCGTT
25221	ጥ አ ር ጥጥ አ ጥጥር	ጥልጥርርጥልሮልሮ	CTGGCAGACG	GCCATGGGCG	TTTGGCAGCA	GIGCIIGGAG
25201	ር አርጥርር እ አርር	TOARCACOT	GCAGAAACTG	CTAAAGCAAA	. ACTTGAAGGA	CCTATGGACG
25443	CCCCCCCAACC	አርርርርጥርርርጥ	CCCCCCCCCAC	CTGGCGGACA	TCATTTTCCC	CGAACGCCIG
25501	COURT A A A CCC	TOCANCAGE	TCTGCCAGAC	TTCACCAGTC	AAAGCATGTT	GCAGAACIII
25561	አርር አአርጥጥ አ	TCCTACACCC	CTCAGGAATC	TTGCCCGCCA	CCTGCTGTGC	ACTICCIAGC
25621	CACTTTCTCC	'	CCGCGAATGC	CCTCCGCCGC	TTTGGGGGCCA	CIGCIACCII
25601	CTCCACCTAC	CCAACTACCT	TGCCTACCAC	TCTGACATAA	TGGAAGACGT	GAGCGGTGAC
25741	CCTCTTACTCC	2 አርጥርጥ ር እርጥር	TCGCTGCAAC	CTATGCACCC	CGCACCGCTC	CCIGGIIIGC
25001	ስ አጥጥር CC ACC	TCCTTAACGA	AAGTCAAATT	ATCGGTACCI	TTGAGCTGCA	GGGTCCCTCG
25061	CCTCACCAA	ACTCCCCGC	TCCGGGGTTG	AAACTCACTC	CGGGGCTGTG	GACGICGGCI
25021	ጥአርርጥጥርርር፤	A ATTTGTACC	TGAGGACTAC	CACGCCCACG	AGATTAGGTT	CTACGAAGAC
25001	CANTCCCCCC	CCCCAAATGC	GGAGCTTACC	: GCCTGCGTC#	A TTACCCAGGG	CCACATTCTT
26041	CCCCAATTCC	A ACCCATCAA	CAAAGCCCGC	CAAGAGTTTC	TGCTACGAAA	GGGACGGGG
26101	CTTTACTTCC	ACCCCCAGTC	CGGCGAGGAG	CTCAACCCA?	A TCCCCCCGCC	GCCGCAGCCC
26161	TATCACCACC	* AGCCGCGGGC	CCTTGCTTCC	: CAGGATGGCA	L CCCAAAAAGA	AGCIGCAGCI
26221	CCCCCCCCC	A CCCACGGACG	AGGAGGAATA	\ CTGGGACAGT	CAGGCAGAGG	AGGTTTTGGA
26201	CCACCACCAC	CACCACATCA	TGGAAGACTG	GGAGAGCCT <i>i</i>	A GACGAGGAAG	CITCUGAGGI
26201	CGAGGAGGA	TCAGACGAAA	CACCGTCACC	CTCGGTCGCA	A TTCCCCTCGC	CGGCGCCCCA
2034.	COMBRAGE	, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,				

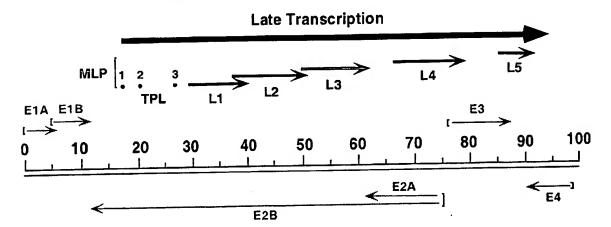
26401	GAAATCGGCA	ACCGGTTCCA	GCATGGCTAC	AACCTCCGCT	CCTCAGGCGC	CGCCGGCACT
		CGACCCAACC				
		CCGTTAGCCC				
26581	GCACAAGAAC	GCCATAGTTG	CTTGCTTGCA	AGACTGTGGG	GGCAACATCT	CCTTCGCCCG
26641	CCGCTTTCTT	CTCTACCATC	ACGGCGTGGC	CTTCCCCCGT	AACATCCTGC	ATTACTACCG
26701	TCATCTCTAC	AGCCCATACT	GCACCGGCGG	CAGCGGCAGC	GGCAGCAACA	GCAGCGGCCA
26761	CACAGAAGCA	AAGGCGACCG	GATAGCAAGA	CTCTGACAAA	GCCCAAGAAA	TCCACAGCGG
		AGGAGGAGGA				
		CAGGATTTTT				
		GAAAATAAAA				
		AGATCAGCTT				
27061	ACTECECECT	GACTCTTAAG	GACTAGTTTC	GCGCCCTTTC	TCAAATTTAA	GCGCGAAAAC
		CCAGCGGCCA				
		CACGCCCTAC				
27101	CTCCCCAACA	CTACTCAACC	CCDATABACT	ACATGAGCGC	GGGACCCCAC	ATGATATCCC
27241	CCCTCAAGA	AATCCGCGCC	CACCGAAACC	CAATTCTCTT	CCAACACCCC	CCTATTACCA
27301	CCACACCOCC	TAATAACCTT	AATCCCCCTA	CTTCCCCCCC	TECCCTECTE	TACCAGGAAA
		CACCACTGTG				
		GCAGCTTGCG				
		GCAGCTTGCG				
2/541	TAACTCACCT	GACAATCAGA	CCCACAMMO	ACAMOCOCCC	CCCCCCCCCC	CCTTCATTCA
27601	CGCTTGGTCT	CCGTCCGGAC	ACTION	COMOCOMOCOMO	TGCCGGCCG1	TOTOCO COCO
2/661	CGCCTCGTCA	GGCAATCCTA	ACTOTGCAGA	TCCCC TCCTC	CONCOUNTRY	CCCDDCCCC
		GCAATTTATT				
		CCACTATCCG				
27841	CGGACGGCTA	CGACTGAATG	TTAAGTGGAG	AGGCAGAGCA	ACTGCGCCTG	AAACACCIGG
27901	TCCACTGTCG	CCGCCACAAG	TGCTTTGCCC	GCGACTCCGG	TGAGTTTTGC	TACTITGAAT
27961	TGCCCGAGGA	TCATATCGAG	GGCCCGGCGC	ACGGCGTCCG	GCTTACCGCC	CAGGGAGAGC
28021	TTGCCCGTAG	CCTGATTCGG	GAGTTTACCC	AGCGCCCCT	GCTAGTTGAG	CGGGACAGGG
28081	GACCCTGTGT	TCTCACTGTG	ATTTGCAACT	GTCCTAACCT	TGGATTACAT	CAAGATCTTT
28141	GTTGCCATCT	CTGTGCTGAG	TATAATAAAT	ACAGAAATTA	AAATATACTG	GGGCTCCTAT
28201	CGCCATCCTG	TAAACGCCAC	CGTCTTCACC	CGCCCAAGCA	AACCAAGGCG	AACCTTACCT
28261	GGTACTTTTA	ACATCTCTCC	CTCTGTGATT	TACAACAGTT	TCAACCCAGA	CGGAGTGAGT
28321	CTACGAGAGA	ACCTCTCCGA	GCTCAGCTAC	TCCATCAGAA	AAAACACCAC	CCTCCTTACC
28381	TGCCGGGAAC	GTACGAGTGC	GTCACCGGCC	GCTGCACCAC	ACCTACCGCC	TGACCGTAAA
28441	CCAGACTTTT	TCCGGACAGA	CCTCAATAAC	TCTGTTTACC	AGAACAGGAG	GTGAGCTTAG
28501	AAAACCCTTA	GGGTATTAGG	CCAAAGGCGC	AGCTACTGTG	GGGTTTATGA	ACAATTCAAG
28561	CAACTCTACG	GGCTATTCTA	ATTCAGGTTT	CTCTAGAATC	GGGGTTGGGG	TTATTCTCTG
28621	TCTTGTGATT	CTCTTTATTC	ጥጥልጥልርጥልልር	COMMOMOMOC	CTAACCCTCC	CCCCCTCCTC
20601			IIMINCIANC	GCTTCTCTGC	CIAMGGCICG	CCGCCIGCIG
7808 ∓	TGTGCACATT	TGCATTTATT				
		TGCATTTATT	GTCAGCTTTT	TAAACGCTGG	GGTCGCCACC	CAAGATGATT
28741	AGGTACATAA	TGCATTTATT TCCTAGGTTT	GTCAGCTTTT ACTCACCCTT	TAAACGCTGG GCGTCAGCCC	GGTCGCCACC ACGGTACCAC	CAAGATGATT CCAAAAGGTG
28741 28801	AGGTACATAA GATTTTAAGG	TGCATTTATT TCCTAGGTTT AGCCAGCCTG	GTCAGCTTTT ACTCACCCTT TAATGTTACA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG	GGTCGCCACC ACGGTACCAC AAGCTAATGA	CAAGATGATT CCAAAAGGTG GTGCACCACT
28741 28801 28861	AGGTACATAA GATTTTAAGG CTTATAAAAT	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC
28741 28801 28861 28921	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT
28741 28801 28861 28921 28981	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC
28741 28801 28861 28921 28981 29041	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC
28741 28801 28861 28921 28981 29041 29101	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT ACTGGCACTT	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA TCTGCTGCAC	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG TGCTATGCTA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC ATTACAGTGC	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG TCGCTTTGGT	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC CTGTACCCTA
28741 28801 28861 28921 28981 29041 29101 29161	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT ACTGGCACTT CTCTATATTA	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA TCTGCTGCAC AATACAAAAAG	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG TGCTATGCTA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC ATTACAGTGC TTTATTGAGG	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG TCGCTTTGGT AAAAGAAAAT	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC CTGTACCCTA GCCTTAATTT
28741 28861 28861 28921 28981 29041 29101 29161 29221	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT ACTGGCACTT CTCTATATTA ACTAAGTTAC	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA TCTGCTGCAC AATACAAAAG AAAGCTAATG	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG TGCTATGCTA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC ATTACAGTGC TTTATTGAGG CTGCTTTACT	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG TCGCTTTGGT AAAAGAAAAT CGCTGCTTCC	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC CTGTACCCTA GCCTTAATTT AAAACAAATT
28741 28861 28921 28981 29041 29101 29161 29221 29281	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT ACTGGCACTT CTCTATATTA ACTAAGTTAC CAAAAAGTTA	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA TCTGCTGCAC AATACAAAAG AAAGCTAATG GCATTATAAT	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG TGCTATGCTA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC ATTACAGTGC TTTATTGAGG CTGCTTTACT TTTAAACCCC	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG TCGCTTTGGT AAAAGAAAAT CGCTGCTTGC CCGGTCATTT	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC CTGTACCCTA GCCTTAATTT AAAACAAATT CCTGCTCAAT
28741 28861 28921 28981 29041 29101 29161 29221 29281 29341	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT ACTGGCACTT CTCTATATTA ACTAAGTTAC CAAAAAGTTA ACCATTCCCC	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA TCTGCTGCAC AATACAAAAG AAAGCTAATG GCATTATAAT TGAACAATTG	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG TGCTATGCTA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC ATTACAGTGC TTTATTGAGG CTGCTTTACT TTTAAACCCC GGATATGCTC	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG TCGCTTTGGT AAAAGAAAAT CGCTGCTTGC CCGGTCATTT CAGCGCTACA	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC CTGTACCCTA GCCTTAATTT AAAACAAATT CCTGCTCAAT ACCTTGAAGT
28741 28801 28861 28921 28981 29041 29101 29161 29221 29281 29341 29401	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT ACTGGCACTT CTCTATATTA ACTAAGTTAC CAAAAAGTTA ACCATTCCCC CAGGCTTCCT	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA TCTGCTGCAC AATACAAAAG AAAGCTAATG GCATTATAAT TGAACAATTG GGATGTCAGC	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG TGCTATGCTA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC ATTACAGTGC TTTATTGAGG CTGCTTTACT TTTAAACCCC GGATATGCTC GGCCAGCACC	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG TCGCTTTGGT AAAAGAAAAT CGCTGCTTGC CCGGTCATTT CAGCGCTACA TGTCCCGCGG	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC CTGTACCCTA GCCTTAATTT AAAACAAATT CCTGCTCAAT ACCTTGAAGT ATTTGTTCCA
28741 28801 28861 28921 28981 29041 29101 29161 29221 29281 29341 29401 29461	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT ACTGGCACTT CTCTATATTA ACTAAGTTAC CAAAAAGTTA ACCATTCCCC CAGGCTTCCT GTCCAACTAC	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA TCTGCTGCAC AATACAAAAG AAAGCTAATG GCATTATAAT TGAACAATTG GGATGTCAGC AGCGACCCAC	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG TGCTATGCTA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC ATTACAGTGC TTTATTGAGG CTGCTTTACT TTTAAACCCC GGATATGCTC GGCCAGCACC ATGACCAACA	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG TCGCTTTGGT AAAAGAAAAT CGCTGCTTGC CCGGTCATTT CAGCGCTACA TGTCCCGCGG CAACCAACGC	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC CTGTACCCTA GCCTTAATTT AAAACAAATT CCTGCTCAAT ACCTTGAAGT ATTTGTTCCA GGCCGCCGCT
28741 28801 28861 28921 28981 29041 29101 29161 29221 29281 29341 29401 29461 29521	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT ACTGGCACTT CTCTATATTA ACTAAGTTAC CAAAAAGTTA ACCATTCCCC CAGGCTTCCT GTCCAACTAC ACCGGACTTA	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA TCTGCTGCAC AATACAAAAG AAAGCTAATG GCATTATAAT TGAACAATTG GGATGTCAGC AGCGACCCAC CATCTACCAC	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG TGCTATGCTA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC ATTACAGTGC TTTATTGAGG CTGCTTTACT TTTAAACCCC GGATATGCTC GGCCAGCACC ATGACCAACA CAAGTTTCTG	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG TCGCTTTGGT AAAAGAAAAT CGCTGCTTGC CCGGTCATTT CAGCGCTACA TGTCCCGCGG CAACCAACGC CCTTTGTCAA	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC CTGTACCCTA GCCTTAATTT AAAACAAATT CCTGCTCAAT ACCTTGAAGT ATTTGTTCCA GGCCGCCGCT TAACTGGGAT
28741 28801 28861 28921 28981 29041 29101 29161 29221 29281 29341 29401 29461 29521 29581	AGGTACATAA GATTTTAAGG CTTATAAAAT AAGTATGCTG TTCCAGGGTA ATTACCATGT ACTGGCACTT CTCTATATTA ACTAAGTTAC CAAAAAGTTA ACCATTCCCC CAGGCTTCCT GTCCAACTAC ACCGGACTTA AACTTGGGCA	TGCATTTATT TCCTAGGTTT AGCCAGCCTG GCACCACAGA TTTATGCTAT AAAGTCATAA ACATGAGCAA TCTGCTGCAC AATACAAAAG AAAGCTAATG GCATTATAAT TGAACAATTG GGATGTCAGC AGCGACCCAC	GTCAGCTTTT ACTCACCCTT TAATGTTACA ACATGAAAAG TTGGCAGCCA AACTTTTATG ACAGTATAAG TGCTATGCTA	TAAACGCTGG GCGTCAGCCC TTCGCAGCTG CTGCTTATTC GGTGACACTA TATACTTTTC TTGTGGCCCC ATTACAGTGC TTTATTGAGG CTGCTTTACT TTTAAACCCC GGATATGCTC GGCCAGCACC ATGACCAACA CAAGTTTCTG CTTATGTTTG	GGTCGCCACC ACGGTACCAC AAGCTAATGA GCCACAAAAA CAGAGTATAA CATTTTATGA CACAAAATTG TCGCTTTGGT AAAAGAAAAT CGCTGCTTGC CCGGTCATTT CAGCGCTACA TGTCCCGCGG CAACCAACGC CCTTTGTCAA TATGCCTTAT	CAAGATGATT CCAAAAGGTG GTGCACCACT CAAAATTGGC TGTTACAGTT AATGTGCGAC TGTGGAAAAC CTGTACCCTA GCCTTAATTT AAAACAAATT CCTGCTCAAT ACCTTGAAGT ATTTGTTCCA GGCCGCCGCT TAACTGGGAT TATTATGTGG

				mmcca ccca c	መርአ አ አር አር አጥ	ርጥጥርጥጥጥርጥ
29701	CTACACCCAA	ACAATGATGG	AATCCATAGA	COCCACOCAC	TOAMACACITE	ACCCTTGTTG
29761	CTTACAGTAT	GATTAAATGA	GACATGATTC	CTCGAGIIII	CATCCAACTA	GACTGCATTC
29821	CGCTTTTTTG	TGCGTGCTCC	ACATIGGCIG	CGGTTTCTCA	CAICGAAGIA	TCCACCCTCA
29881	CAGCCTTCAC	AGTCTATTTG	CTTTACGGAT	TIGICACCCI	CHCGCTCATC	TOCAGCCION
29941	TCACTGTGGT	CATCGCCTTT	ATCCAGTGCA	TTGACTGGGT	COUNCEMBER	አጥጥርጥጥጥልልጥ
30001	TCAGACACCA	TCCCCAGTAC	AGGGACAGGA	CTATAGCTGA	CENTROCCCC	ተመጥርጥጥር CCC
30061	TATGAAATTT	ACTGTGACTT	TTCTGCTGAT	TATTTGCACC	CIMICIGCGI	ATATTCCAAG
30121	GACCTCCAAG	CCTCAAAGAC	ATATATCATG	DAGATICACI	MANCCA ANCA	ጥርጥርጥርጥጥልጥ
30181	TTGCTACAAT	GAAAAAAGCG	ATCTTTCCGA	AGCCTGGTTA	TAIGCAAICA	TTCCCTGGAA
30241	GGTGTTCTGC	AGTACCATCT	TAGCCCTAGC	TATATATCCC	CCTATCCTTC	CACTGCAACA
30301	ACGAATAGAT	GCCATGAACC	ACCCAACTIT	TOTAL COORDECC	CCCACTTCTC	CCACCCCCAC
30361	AGTTGTTGCC	GGCGGCTTTG	TCCCAGCCAA	TCAGCCTCGC	CACCCTICIC	CTACAAATGG
30421	TGAAATCAGC	TACTTTAATC	TAACAGGAGG	AGATGACTGA	CACCCIAGAI	CACCAACAGC
30481	ACGGAATTAT	TACAGAGCAG	CGCCTGCTAG	AAAGACGCAG	CUCCAAAACC	CCTATCTTT
30541	GCATGAATCA	AGAGCTCCAA	GACATGGTTA	ACTIGUACUA	GIGCAAAAGG	CCCCTTACCT
30601	GTCTGGTAAA	GCAGGCCAAA	GTCACCTACG	ACAGTAATAC	CACCACAAAAC	CCCATTACCA
30661	ACAAGTTGCC	AACCAAGCGT	CAGAAATTGG	TGGTCATGGT	ACCUMCTO A	CCATTACCA
30721	TAACTCAGCA	CTCGGTAGAA	ACCGAAGGCT	GCATTCACTC	MCCTTGTCAA	መመጥ አርጥል ልጥ
30781	ATCTCTGCAC	CCTTATTAAG	ACCCTGTGCG	GTCTCAAAGA	CONNUMERO	CTCCACTTA
30841	TAAAAAAAA	AATAAAGCAT	CACTTACTTA	AAATCAGTTA	CONCOMMENT	CCTCCCTCCA
30901	TTCAGCAGCA	CCTCCTTGCC	CTCCTCCCAG	CTCTGGTATT	COMPCONCOC	AMCCCCACCC
30961	AACTTTCTCC	ACAATCTAAA	TGGAATGTCA	GTTTCCTCCT	GTTCCTGTCC	CTTCAACCCC
31021	ACTATCTTCA	TGTTGTTGCA	GATGAAGCGC	GCAAGACCGT	CTGAAGATAC	TO TO TO TO THE TOTAL THE TOTAL TO THE TOTAL TOTAL TO THE
31081	GTGTATCCAT	ATGACACGGA	AACCGGTCCT	CCAACTGTGC	CTTTTCTTAC	CCTCCCTTT
31141	GTATCCCCCA	ATGGGTTTCA	AGAGAGTCCC	CCTGGGGTAC	CONNECCE	CULVICCOV
31201	CCTCTAGTTA	CCTCCAATGG	CATGCTTGCG	CTCAAAATGG	GCAACGGCCT	CAAAAAAACC
31261	GAGGCCGGCA	ACCTTACCTC	CCAAAATGTA	ACCACTGTGA	GCCCACCTCT	ACCCCUDA ACT
31321	AAGTCAAACA	TAAACCTGGA	AATATCTGCA	CCCCTCACAG	TTACCTCAGA	AGCCCIAACI
31381	GTGGCTGCCG	CCGCACCTCT	AATGGTCGCG	GGCAACACAC	TUACCATGCA	ATCACAGGCC
31441	CCGCTAACCG	TGCACGACTC	CAAACTTAGC	ATTGCCACCC	AAGGACCCCT	CACAGIGICA
31501	GAAGGAAAGC	TAGCCCTGCA	AACATCAGGC	CCCCTCACCA	CCACCGATAG	TCACIACCCII
31561	ACTATCACTG	CCTCACCCCC	TCTAACTACT	GCCACTGGTA	CCTTGGGCAT	TGACTIGAAA
31621	GAGCCCATTT	ATACACAAAA	TGGAAAACTA	GGACTAAAGT	ACGGGGCTCC	TTTGCATGIA
31681	ACAGACGACC	TAAACACTTT	GACCGTAGCA	ACTGGTCCAG	GIGIGACIAI	TAATAATACT
31741	TCCTTGCAAA	CTAAAGTTAC	TGGAGCCTTG	GGTTTTGATT	CACAAGGCAA	TAIGCAACII
31801	AATGTAGCAG	GAGGACTAAG	GATTGATTCT	CAAAACAGAC	GCCTTATACT	TGATGITAGI
21061	ጥአጥርርርጥጥጥር	አጥርርጥር AAAA	CCAACTAAAT	CTAAGACTAG	GACAGGGCCC	TCTTTTTATA
31921	AACTCAGCCC	ACAACTTGGA	TATTAACTAC	AACAAAGGCC	TTTACTTGTT	TACAGCTICA
31981	AACAATTCCA	AAAAGCTTGA	GGTTAACCTA	AGCACTGCCA	AGGGGTTGAT	GTTTGACGCT
32041	ACAGCCATAG	CCATTAATGC	AGGAGATGGG	CTTGAATTTG	GTTCACCTAA	TGCACCAAAC
32101	ACAAATCCCC	TCAAAACAAA	AATTGGCCAT	GGCCTAGAAT	TTGATTCAAA	CAAGGCTATG
32161	GTTCCTAAAC	TAGGAACTGG	CCTTAGTTTT	GACAGCACAG	GTGCCATTAC	AGTAGGAAAC
32221	AAAAATAATC	ATAAGCTAAC	TTTGTGGACC	ACACCAGCTC	CATCTCCTAA	CIGIAGACIA
32281	AATGCAGAGA	AAGATGCTAA	ACTCACTTTG	GTCTTAACAA	AATGTGGCAG	TCAAATACTT
22241		· ሶልሮጥጥጥጥርርር	TGTTAAAGGC	AGTTTGGCTC	CAATATCIGG	MACAGITCAM
32401	ል ርጥር ርጥር ልጥር	' TTATTATAAG	ATTTGACGAA	AATGGAGTGC	TACTAAACAA	TICCTICCIG
22461	CACCCAGAAT	NTTGGAACTT	TAGAAATGGA	GATCTTACT	AAGGCACAGC	CTATACAAAC
22521	CCTCTTCCAT	TTATCCCTAA	CCTATCAGCT	TATCCAAAAT	CTCACGGTAA	AACTGCCAAA
22501	አርሞአአርአጥጥር	TOACTOAACT	TTACTTAAAC	GGAGACAAAA	CTAAACCTGT	AACACTAACC
22641	3 mm 3 C 3 C m 3 7	ACCCTACACA	CCAAACACGA	GACACAACTO	CAAGTGCATA	CTCTATGTCA
22201	መመመመር አመርርር	 * * * * * * * * * * * * * * * * * * *	CCACAACTAC	ATTAATGAAA	TATTTGCCAC	ATCCTCTTAC
22761	አ ር መውመጥጥጥር አ ባ	' አርስጥጥርርርር	AGAATAAAGA	ATCGTTTGT	TTATGTTTCA	ACGIGITIAL
22921	ጥጥጥጥር ል ልጥጥር	CACAAAATTT	CAAGTCATTI	' TTCATTCAGT	· AGTATAGCCC	CACCACCACA
22001	መክርር መመከመስ	` አርልጥር እርርር ጥ	ACCTTAATCA	. AACTCACAGA	ACCCTAGTAT	TUAACCTGCC
32941	ACCTCCCTCC	CAACACACAG	AGTACACAGT	CCTTTCTCC	CGGCTGGCCI	TAAAAAGCAT

33001	CATATCATGG	GTAACAGACA	TATTCTTAGG	TGTTATATTC	CACACGGTTT	CCTGTCGAGC
33061	CAAACGCTCA	TCAGTGATAT	TAATAAACTC	CCCGGGCAGC	TCACTTAAGT	TCATGTCGCT
33121	GTCCAGCTGC	TGAGCCACAG	GCTGCTGTCC	AACTTGCGGT	TGCTTAACGG	GCGGCGAAGG
33181	AGAAGTCCAC	GCCTACATGG	GGGTAGAGTC	ATAATCGTGC	ATCAGGATAG	GGCGGTGGTG
33241	CTGCAGCAGC	GCGCGAATAA	ACTGCTGCCG	CCGCCGCTCC	GTCCTGCAGG	AATACAACAT
33301	GGCAGTGGTC	TCCTCAGCGA	TGATTCGCAC	CGCCCGCAGC	ATAAGGCGCC	TTGTCCTCCG
					TAACTGCAGC	
33421	AATATTGTTC	AAAATCCCAC	AGTGCAAGGC	GCTGTATCCA	AAGCTCATGG	CGGGGACCAC
33481	AGAACCCACG	TGGCCATCAT	ACCACAAGCG	CAGGTAGATT	AAGTGGCGAC	CCCTCATAAA
33541	CACGCTGGAC	ATAAACATTA	CCTCTTTTGG	CATGTTGTAA	TTCACCACCT	CCCGGTACCA
33601	TATAAACCTC	TGATTAAACA	TGGCGCCATC	CACCACCATC	CTAAACCAGC	TGGCCAAAAC
					CAATGACAGT	
					ATGTTGGCAC	
					GTTAGAACCA	
33841	AACAACCCAT	TCCTGAATCA	GCGTAAATCC	CACACTGCAG	GGAAGACCTC	GCACGTAACT
33901	CACGTTGTGC	ATTGTCAAAG	TGTTACATTC	GGGCAGCAGC	GGATGATCCT	CCAGTATGGT
33961	AGCGCGGGTT	TCTGTCTCAA	AAGGAGGTAG	ACGATCCCTA	CTGTACGGAG	TGCGCCGAGA
					ACGCCGGACG	
					GCGTCTCCGG	
34141	TAGATCGCTC	TGTGTAGTAG	TTGTAGTATA	TCCACTCTCT	CAAAGCATCC	AGGCGCCCCC
34201	TGGCTTCGGG	TTCTATGTAA	ACTCCTTCAT	GCGCCGCTGC	CCTGATAACA	TCCACCACCG
34261	CAGAATAAGC	CACACCCAGC	CAACCTACAC	ATTCGTTCTG	CGAGTCACAC	ACGGGAGGAG
					AAAGATTATC	
34381	AAATGAAGAT	CTATTAAGTG	AACGCGCTCC	CCTCCGGTGG	CGTGGTCAAA	CTCTACAGCC
					CTTCCAAAAG	
34501	CTCACGTCCA	AGTGGACGTA	AAGGCTAAAC	CCTTCAGGGT	GAATCTCCTC	TATAAACATT
					ACCTTCTCAA	
					GCTCCAGAGC	
					TTCCTCACAG	
					TAGGTCCCTT	
					CACTTCCCCG	
					AGCTATGCTA	
					GCAAGGTGCT	
34981	TCAGGCAAAG	CCTCGCGCAA	AAAAGAAAGC	ACATCGTAGT	CATGCTCATG	CAGATAAAGG
					TTCTCTCAAA	
					TTAAACATTA	
					TACGGCCATG	
					CAGCTCCTCG	
35281	GAGTCATAAT	GTAAGACTCG	GTAAACACAT	CAGGTTGATT	CATCGGTCAG	TGCTAAAAAG
					GAGACAACAT	
					AAACACCTGA	
					ACAGCGCTTC	
35521	CCTAACAGTC	AGCCTTACCA	GTAAAAAAGA	AAACCTATTA	AAAAAACACC	ACTCGACACG
					GCAGAGCGAG	
					CCAGAAAACC	
					CAAATCGTCA	
					CCCAACACAT	
					CGCGCCACGT	
35881	ACCCCCTCAT	TATCATATTG	GCTTCAATCC	AAAATAAGGT	ATATTATTGA	TGATG

68/92

Structure of the Ad6 Genome



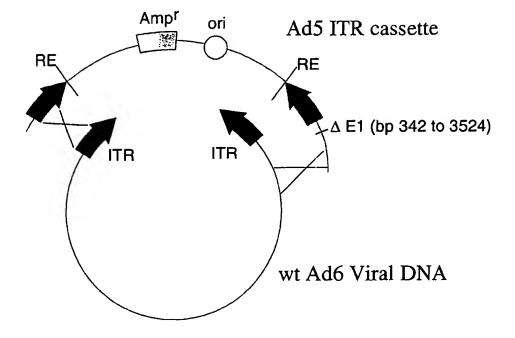


FIG. 10

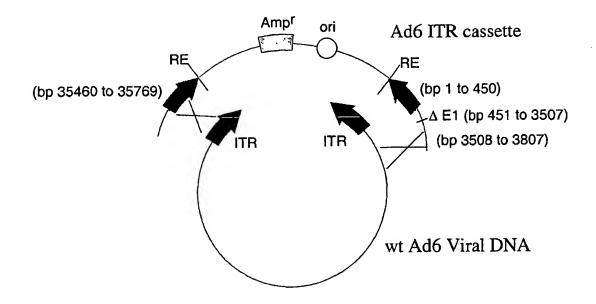
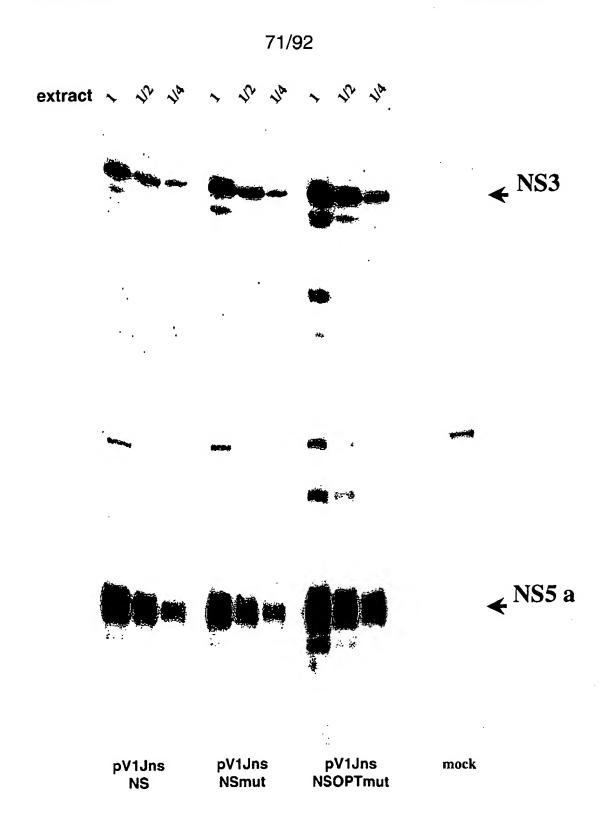


FIG. 11



Western blot on whole-cell extracts from 293 cells transfected with plasmid DNA expressing the different HCV NS cassettes. Mature NS3 and NS5A products were detected with specific antibodies.

FIG. 12

WO 03/031588

72/92

					Pep pool				
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)	DMSC
	#31	41	135	19	44	25	17	137	8
	#32	121	783	7 7	144	13	22	604	4
	#33	8	32	3	11	6	6	43	3
	#34	16	139	13	47	31	25	151	2
-V1: NC	#35	21	101	40	32	21	20	75	1
pV1jns-NS	#36	18	26	24	25	5	7	29	6
	#37	19	73	15	39	8	20	49	2
	#38	133	575	74	345	75	63	515	5
	#39	40	183	10	85	14	9	148	2
	#40	66	465	29	111	15	16	189	0
	Geomean	33	146	21	57	15	16	123	na

				Pep pool				
mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)	DMSC
#41	39	293	58	187	5	4	248	1
#42	21	220	46	107	26	10	189	4
#43	76	134	12	78	8	6	144	2
#44	30	45	20	52	4	8	40	4
#45	36	100	17	56	4	6	116	3
#46	67	172	16	138	8	9	145	3
#47	34	131	28	38	9	5	118	1
#48	55	316	43	107	9	7	277	5
#49	6	131	5	25	4	1	91	0
#50	13	93	11	11	5	1	76	1
Geomean	30	142	20	61	7	5	126	na

pV1jns-NSmut

	1				Pep pool				
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)	DMSC
	#51	53	409	34	84	11	25	271	4
	#52	140	660	65	276	23	36	377	2
	#53	58	553	48	105	23	18	564	1
	#54	50	105	35	134	10	16	80	2
V1jns-NSOPTmut	#55	14	80	11	35	4	7	91	6
•	#56	14	342	30	101	23	14	207	1
	#57	63	325	66	239	17	24	123	1
	#58	75	542	66	168	127	93	191	0
	#59	65	468	40	124	18	23	344	4
	#60	27	142	48	16	7	В	77	0
	Geomean	45	295	40	99	16	20	188	na

IFNY ELIspot on splenocytes from C57black6 mice immunized with two injections of 25µg DNA/dose with GET of plasmid vectors expressing the different HCV NS cassettes. Data are expressed as SFC/106 PBMC.

FIG. 13A

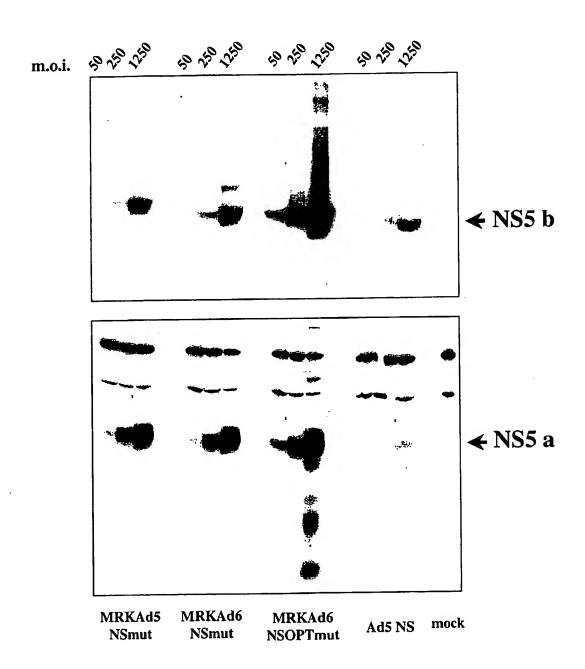
73/92

				Pe	p pool			
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
	#51	219	699	634	486	487	264	34
	#52	67	302	347	167	111	87	9
	#53	59	460	400	246	244	136	26
	#54	139	817	685	236	547	223	24
	#55	96	904	542	277	256	337	17
pV1jns-NS	#56	225	603	686	156	350	240	56
	#57	44	288	211	148	100	141	4
	#58	37	262	221	53	58	62	3
	#59	131	975	928	159	305	284	14
	#60	93	475	464	77	206	113	12
	geo mean	111	579	512	201	266	189	20
				Pe	p pool			
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
	#61	72	840	515	219	278	249	19
	#62	294	1881	1266	365	434	411	63
	#63	73	415	422	103	141	99	41
**** ***	#64	66	824	486	175	162	144	18
pV1jns-NSmut	#66	24	313	168	53	47	42	. 5
	#67	15	230	253	94	25	39	2
	#68	53	354	252	89	101	86	15
	#69	271	895	909	518	322	285	74
	#70	417	1303	1186	468	557	267	34
	geo mean	143	784	606	232	230	180	30
				Pe	p pool			
	mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
	#71	206	944	890	342	207	397	47
	#72	393	1655	1151	575	626	401	72
	#73	123	522	515	319	223	198	21
//41 NOOPM .	#74	500	1414	1419	878	1035	1122	137
V1jns-NSOPTmut	#75	286	812	873	382	543	267	31
	#76	224	1143	942	218	420	281	22
	#77	95	643	630	169	385	218	15
	#78	401	1302	1068	538	608	623	12
	#79	108	1190	914	199	265	215	4
	#80	122	511	546	189	286	190	13
	geo mean	209	941	854	331	406	329	24

IFNy ELIspot on splenocytes from BalbC mice immunized with two injections of 50µg DNA/dose with GET of plasmid vectors expressing the different HCV NS cassettes. Data are expressed as SFC/10⁶ PBMC.

FIG. 13B

74/92



Western blot on whole-cell extracts from HeLa cells infected at different multiplicity of infection (m.o.i.; indicated at the top) with Adenovectors expressing the different HCV NS cassettes. Mature NS5B and NS5A products were detected with specific antibodies.

FIG. 14

75/92

				Pep pool			
mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8	ep)DMSO
#1	14	492	9	27	10	554	7
#2	8	440	2	26	5	438	0
#3	12	92	5	12	7	73	4
#4	16	388	6	40	6	228	2
#6	8	210	4	31	3	238	3
#7	7	133	13	16	0	128	9
#8	11	342	25	55	22	267	12
#9	5	345	0	45	5	285	3
#10	22	888	3	65	25	799	1
Geomean	10	305	na	31	na	269	na

Pep pool I(NS5a) L+M(NS35b) 1480(CD8 ep)DMSO F(NS3p) G(NS3h) H(NS4) mouse #11 #12 #13 #14 #15 #16 #17 #18 #19

na

MRKAd5-NSmut

Ad5-NS

				Pep pool			
mouse	F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8	ep)DMSO
#21	6	584	5	27	4	491	2
#22	6	231	3	12	3	235	0
#23	8	482	1	18	1	511	0
#24	14	1120	6	38	10	1004	5
#25	1	311	3	9	0	382	1
#26	29	903	3	60	5	751	5
#27	35	1573	4	40	4	1277	4
#28	7	406	5	15	1	443	3
#29	4	461	3	12	3	515	_ 3
Geomean	8	567	3	21	na	554	na

MRKAd6-NSmut

IFNy ELISPOT on splenocytes from C57black6 mice immunized with two injections of 109 vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/106 PBMC.

Geomean

76/92

	Ad5	-NS 10 ¹⁰ vp/d	ose
Pep pools	96074	134T	063Q
F (NS3p)	374	11	74
G (NS3h)	359	1070	1455
H (NS4)	376	30	64
I (NS5a)	240	40	63
L (NS5b)	226	29	121
M (NS5b)	511	23	35
DMSO	128	3	31

	MRK Ad6-NSmut 10 ¹⁰ vp/dose						
Pep pools	S207	035Q	057Q				
F (NS3p)	363	382	150				
G (NS3h)	180	316	119				
H (NS4)	126	113	62				
I (NS5a)	1780	688	114				
L (NS5b)	447	111	81				
M (NS5b)	153	38	16				
DMSO	9	6	9				

IFNY ELISPOT on PBMC from Rhesus monkeys immunized with one injection of 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16A

77/92

	MRK Ad5-NSmut 10 ¹⁰ vp/dose					
Pep pools	S201	075 <u>Q</u>	137Q			
F (NS3p)	928	69	254			
G(NS3h)	317	436	98			
H (NS4)	56	101	45			
I (NS5a)	1530	1100	413			
L (NS5b)	149	23	92			
M (NS5b)	398	32	80			
DMSO	29	6	29			

	MRK Ad6-NSOPTmut 10 ¹⁰ vp/dose					
Pep pools	98D209	106Q	113Q			
F (NS3p)	3110	263	404			
G(NS3h)	2115	642	1008			
H (NS4)	373	72	19			
I (NS5a)	103	37	347			
L (NS5b)	149	22	10			
M (NSSb)	314	428	- 19			
DMSO	0	1	3			

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with one injection of 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/106 PBMC.

FIG. 16B

78/92

	Ad5-NS 10 ¹¹ vp/dose						
Pep pools	99C008	97N104	97X008	99C026			
F (NS3p)	28	1026	579	889			
G (NS3h)	1279	188	103	2453			
H (NS4)	18	39	138	109			
I (NS5a)	131	1068	172	141			
L (NS5b)	78	144	103	32			
M (NS5b)	24	68	47	84			
DMSO	3	16	1	19			

	MRKAd6-NSmut 10 ¹¹ vp/dose							
Pep pools	98C047 97C055 93G 977							
F (NS3p)	477	25	93	1022				
G(NS3h)	959	398	81	1513				
H (NS4)	36	14	99	53				
1 (NS5a)	171	45	1237	98				
L (NS5b)	18	32	23	51				
M (NS5b)	88	4	13	40				
DMSO	8	3	1	5				

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/106 PBMC.

FIG. 16C

79/92

	MRKAd5-NSmut 10 ¹¹ vp/dose						
Pep pools	99C059	99C060	97X009	96069			
F (NS3p)	28	81	1308	1618			
G (NS3h)	2600	161	1008	123			
H (NS4)	31	74	101	40			
1 (NS5a)	181	99	69	96			
L (NS5b)	24	31	40	20			
M (NS5b)	11	58	38	164			
DMSO	6	15	1	16			

IFNy ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

80/92

	MRK Ad5-NSmut 10 10 vp/dose					
Pep pools	S201	075Q	137 <u>Q</u>			
pool F (NS3p)	881	1755	73			
pool G (NS3h)	573		•			
pool H (NS4)		3541				
pool I (NS5a)	2094		39			
pool L (NS5b)						
pool M (NS5b)	756					
DMSO	319	117	44			

	MRK Ad6-N	10 vp/dose	
Pep pools	98D209	106Q	113 <u>Q</u>
pool F (NS3p)	5073	84	952
pool G (NS3h)	2376	160	3325
pool H (NS4)	700		
pool I (NS5a)			1106
pool L (NS5b)			
pool M (NS5b)	530	706	
DMSO	43	47	28

	MRK Ad	6-NSmut 10	¹⁰ vp/dose
Pep pools	S207	035Q	057 <u>Q</u>
pool F (NS3p)	118	480	
pool G (NS3h)		196	
pool H (NS4)			
pool I (NS5a)	3340	933	
pool L (NS5b)	118		
pool M (NS5b)			
DMSO .	145	34	

IFN γ ICS on PBMC from Rhesus monkeys immunized with two injections at four weeks interval with 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as number of positive IFN γ /CD3/CD8 per 10^6 lymphocytes.

FIG. 17A

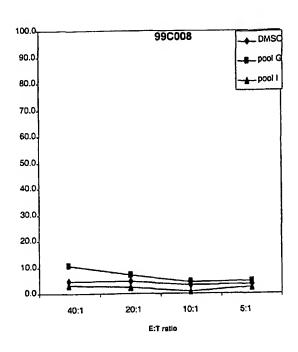
81/92

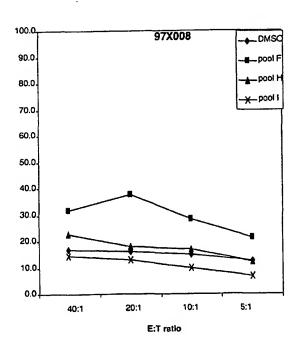
		Δ	Ad5-NS 10	11 vp/do	se		
	Pep pools	99C008	97N104	97X008	99C026		
_	F (NS3p)		1703	1136	615		
	G (NS3h)	3153			2787		
	H (NS4)						
	I (NS5a)		2233				
	L (NS5b)						
	M (NS5b)						
	DMSO	125	98	130	0		
		MRKA	Ad6-NSmi	ut 10 ¹¹ v	p/dose		
	Pep pools	98C047	97C055	93G	97X014		
	F (NS3p)	1024			948		
	G(NS3h)	3246	353		1074		
	H (NS4)			316			
	I (NS5a)			6224			
	L (NS5b)						
	M (NS5b)						
	DMSO	49	23	37	93		
		MRKAd5-NSmut 10 11 vp/dose					
	Pep pools	99C059	99C060	97X009	96069		
	F (NS3p)			2266	5053		
	G (NS3h)	2434	316	1018			
	H (NS4)						
	I (NS5a)	1					
	L (NS5b)						
	M (NS5b)		_		205		
	DMSO	13	110	119	15		
	200						

IFNY ICS on PBMC from Rhesus monkeys immunized with two injections at four weeks interval with 10¹¹ vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as number of positive IFNY/CD3/CD8 per 10⁶ lymphocytes.

FIG. 17B



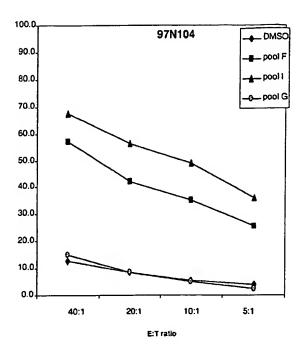


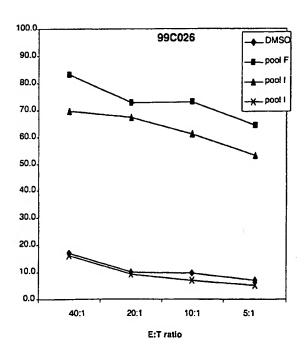


Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10¹¹vp/dose of Ad5-NS.

FIG. 18A



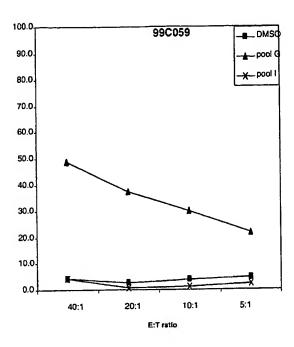


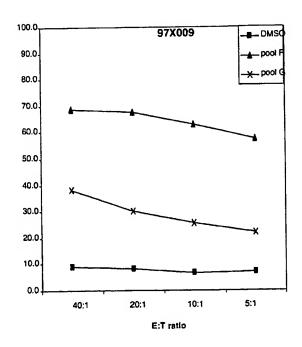


Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10¹¹vp/dose of Ad5-NS.

FIG. 18B

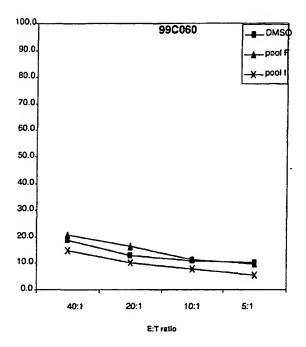
84/92

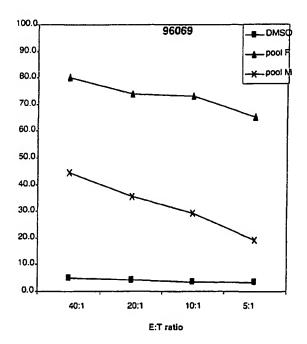




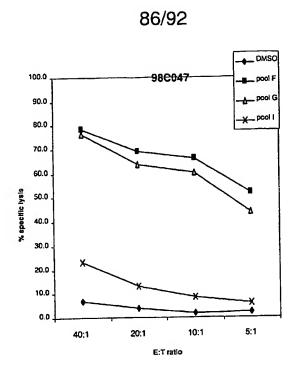
Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 1011vp/dose of MRKAd5-NSmut.

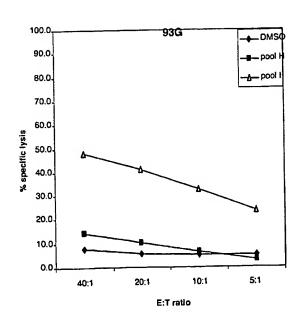






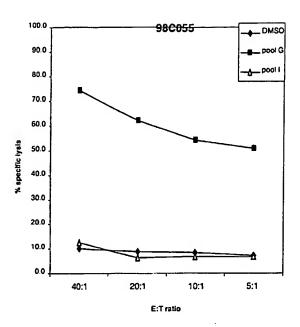
Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 1011vp/dose of MRKAd5-NSmut

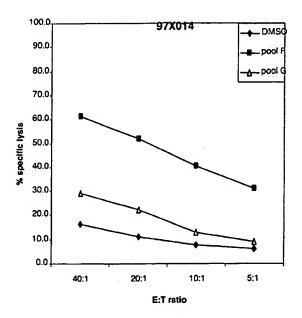




Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 1011vp/dose of MRKAd6-NSmut.







Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10¹¹vp/dose of MRKAd6-NSmut.

FIG. 18F

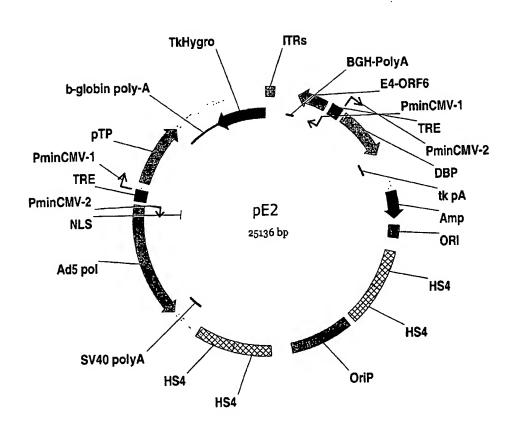


FIG. 19

1	GCCACCATGG	CCCCCATCAC	CGCCTACAGC	CAGCAGACCA	GGGGCCTGCT
51	GGGCTGCATC	ATCACCAGCC	TGACCGGACG	CGACAAGAAC	CAGGTGGAGG
101	GAGAGGTGCA	GGTGGTGAGC	ACCGCTACCC	AGAGCTTCCT	GGCCACCTGC
151	GTGAACGGCG	TGTGCTGGAC	CGTGTACCAC	GGAGCCGGAA	GCAAGACCCT
201	GGCCGGACCC	AAGGCCCTA	TCACCCAGAT	GTACACCAAT	GTGGATCAGG
251	ATCTGGTGGG	CTGGCAGGCC	CCTCCCGGAG	CCAGGAGCCT	GACACCCTGT
301	ACCTGTGGAA	GCAGCGACCT	GTACCTGGTG	ACACGCCACG	CCGATGTGAT
351	CCCCGTGAGG	CGCAGGGGCG	ATTCTCGCGG	AAGCCTGCTG	AGCCCTAGGC
401	CCGTGAGCTA	CCTGAAGGGC	AGCAGCGGAG	GACCCCTGCT	GTGTCCTTCT
451	GGCCATGCCG	TGGGCATTTT	TCGCGCTGCC	GTGTGTACCA	GGGGCGTGGC
501	CAAAGCCGTG	GATTTTGTGC	CCGTGGAAAG	CATGGAGACC	ACCATGCGCA
551	GCCCTGTGTT	CACCGACAAC	AGCTCTCCCC	CTGCCGTGCC	CCAATCATTC
601	CAGGTGGCTC	ACCTGCACGC	CCCTACCGGA	TCTGGCAAGA	GCACCAAGGT
651	GCCCGCTGCC	TACGCCGCTC	AGGGCTACAA	GGTGCTGGTG	CTGAACCCCA
701	GCGTGGCCGC	TACCCTGGGC	TTCGGCGCTT	ACATGAGCAA	GGCCCATGGC
751	ATCGACCCCA	ACATCCGCAC	AGGCGTGCGC	ACCATCACCA	CCGGAGCTCC
801	CGTGACCTAC	AGCACCTACG	GCAAGTTCCT	GGCCGATGGA	GGCTGCAGCG
851	GAGGAGCCTA	CGACATCATC	ATCTGCGACG	AGTGCCACAG	CACCGACAGC
901	ACCACCATCC	TGGGCATTGG	CACCGTGCTG	GATCAGGCCG	AAACAGCTGG
951	AGCCAGGCTG	GTGGTGCTGG	CCACAGCTAC	CCCTCCTGGC	AGCGTGACCG
1001	TGCCCCATCC	CAATATCGAG	GAGGTGGCCC	TGAGCAACAC	AGGCGAGATC
1051	CCCTTCTACG	GCAAGGCCAT	CCCCATCGAG	GCCATCCGCG	GAGGCAGGCA
1101	CCTGATCTTC	TGCCACAGCA	AGAAGAAGTG	CGACGAGCTG	GCTGCCAAGC
1151	TGAGCGGACT	GGGCATCAAC	GCCGTGGCCT	ACTACAGGGG	CCTGGACGTG
1201	TCAGTGATCC	CCACCATCGG	CGATGTGGTG	GTGGTGGCCA	CCGACGCCCT
1251	GATGACAGGC	TACACCGGAG	ACTTCGACAG	CGTGATCGAC	TGCAACACCT
1301	GCGTGACCCA	GACCGTGGAC	TTCAGCCTGG	ACCCCACCTT	CACCATCGAA
1351	ACCACCACCG	TGCCTCAGGA	TGCTGTGAGC	AGGAGCCAGA	GGCGCGGACG
1401	CACCGGAAGG	GGCAGGCGCG	GAATTTATCG	CTTTGTGACC	CCTGGCGAAA
1451	GGCCCTCTGG	CATGTTCGAC	AGCAGCGTGC	TGTGCGAGTG	CTACGACGCT
1501	GGCTGCGCTT	GGTACGAGCT	GACACCCGCT	GAAACCAGCG	TGCGCCTGCG
1551	CGCTTATCTG	AATACCCCTG	GCCTGCCCGT	GTGTCAGGAC	CACCTGGAGT

FIG. 20A

1601		CGTGTTCACA			
1651	AGCCAGACCA	AGCAGGCTGG	CGACAACTTC	CCCTATCTGG	TGGCCTATCA
1701	GGCCACCGTG	TGTGCTAGGG	CCCAAGCTCC	ACCTCCTTCA	TGGGACCAGA
1751	TGTGGAAGTG	CCTGATCCGC	CTGAAGCCCA	CCCTGCACGG	CCCTACCCCT
1801	CTGCTGTACC	GCCTGGGAGC	CGTGCAGAAC	GAGGTGACCC	TGACCCACCC
1851	CATCACCAAG	TACATCATGG	CCTGCATGAG	CGCTGATCTG	GAAGTGGTGA
1901	CCAGCACCTG	GGTGCTGGTG	GGAGGCGTGC	TGGCCGCTCT	GGCTGCCTAC
1951	TGCCTGACCA	CCGGAAGCGT	GGTGATCGTG	GGACGCATCA	TCCTGAGCGG
2001	AAGGCCCGCT	ATCGTGCCCG	ATCGCGAGTT	CCTGTACCAG	GAGTTCGACG
2051	AGATGGAGGA	GTGTGCCAGC	CACCTGCCCT	ACATCGAGCA	GGGCATGCAG
2101	CTGGCCGAAC	AGTTCAAGCA	GAAGGCCCTG	GGCCTGCTGC	AGACAGCCAC
2151	CAAACAGGCC	GAAGCTGCCG	CTCCCGTGGT	GGAAAGCAAG	TGGAGGGCCC
2201	TGGAGACCTT	CTGGGCTAAG	CACATGTGGA	ACTTCATCTC	TGGCATCCAG
2251	TACCTGGCCG	GACTGAGCAC	CCTGCCTGGC	AACCCCGCTA	TCGCCAGCCT
2301	GATGGCCTTC	ACCGCTAGCA	TCACCTCTCC	CCTGACCACC	CAGAGCACCC
2351	TGCTGTTCAA	CATTCTGGGC	GGATGGGTGG	CCGCTCAGCT	GGCCCCTCCT
2401	TCAGCTGCTT	CTGCCTTTGT	GGGCGCTGGC	ATTGCCGGAG	CCGCTGTGGG
2451	CAGCATTGGC	CTGGGCAAAG	TGCTGGTGGA	TATTCTGGCT	GGCTATGGCG
2501	CTGGCGTGGC	CGGAGCCCTG	GTGGCCTTCA	AGGTGATGAG	CGGAGAGATG
2551	CCCAGCACCG	AGGACCTGGT	GAACCTGCTG	CCTGCCATTC	TGAGCCCTGG
2601	AGCCCTGGTG	GTGGGCGTGG	TGTGTGCTGC	CATTCTGAGG	CGCCATGTGG
2651	GACCCGGAGA	GGGCGCTGTG	CAGTGGATGA	ACCGCCTGAT	CGCCTTCGCC
2701	TCTCGCGGAA	ACCACGTGAG	CCCTACCCAC	TACGTGCCTG	AGAGCGACGC
2751	CGCTGCCAGG	GTGACCCAGA	TCCTGAGCAG	CCTGACCATC	ACCCAGCTGC
2801	TGAAGCGCCT	GCACCAGTGG	ATCAACGAGG	ACTGCAGCAC	ACCCTGCAGC
2851	GGAAGCTGGC	TGAGGGACGT	GTGGGACTGG	ATCTGCACC	TGCTGACCGA
2901	CTTCAAGACO	TGGCTGCAGA	GCAAGCTGCT	GCCCCAACT	CCTGGCGTGC
2951	CCTTCTTCTC	ATGCCAGCGC	GGATACAAG	GCGTGTGGAC	GGGCGATGGC
3001	ATCATGCAGA	CCACCTGTCC	CTGCGGAGCC	CAGATCACAC	GCCACGTGAA
3051	GAACGGCAG	ATGCGCATCG	TGGGCCCTA	A GACCTGCAG	AACACCTGGC
3101	ACGGCACCT	CCCCATCAAC	GCCTACACCA	A CCGGACCCT	G CACACCCAGC
3151	CCTGCTCCC	A ACTACAGCAG	GGCCCTGTG	G AGGGTGGCT	G CCGAGGAGTA

2221	000003 0000	N O C N C C C C C C C C C C C C C C C C	CACACOOOCCA	CHYCCHCYCC	CC3 3 TC3 CC3
3201					GGAATGACCA
3251	CCGACAACGT	GAAGTGTCCC	TGTCAGGTGC	CCGCTCCCGA	ATTTTTTACC
3301	GAAGTGGATG	GCGTGCGCCT	GCATCGCTAT	GCCCCTGCCT	GTAGGCCCCT
3351	GCTGCGCGAA	GAAGTGACCT	TCCAGGTGGG	CCTGAACCAG	TACCTGGTGG
3401	GCAGCCAGCT	GCCCTGCGAG	CCTGAGCCCG	ATGTGGCCGT	GCTGACCAGC
3451	ATGCTGACCG	ACCCCAGCCA	CATCACAGCC	GAAACCGCTA	AAAGGCGCCT
3501	GGCCAGGGGC	TCTCCTCCAA	GCCTGGCCTC	AAGCAGCGCT	AGCCAGCTGT
3551	CTGCTCCCAG	CCTGAAGGCC	ACCTGCACCA	CCCACCACGT	GAGCCCCGAC
3601	GCCGACCTGA	TCGAGGCCAA	CCTGCTGTGG	CGCCAGGAGA	TGGGCGGCAA
3651	CATCACCCGC	GTGGAGAGCG	AGAACAAGGT	GGTGGTGCTG	GACAGCTTCG
3701	ACCCCCTGCG	CGCCGAGGAG	GACGAGCGCG	AGGTGAGCGT	GCCCGCCGAG
3751	ATCCTGCGCA	AGAGCAAGAA	GTTCCCCGCT	GCCATGCCCA	TCTGGGCTAG
3801	ACCTGATTAC	AACCCTCCCC	TGCTGGAGAG	CTGGAAGGAC	CCTGATTACG
3851	TGCCTCCAGT	GGTGCATGGC	TGTCCTCTGC	CTCCCATTAA	AGCCCCTCCT
3901	ATTCCACCTC	CTAGGCGCAA	AAGGACCGTG	GTGCTGACAG	AAAGCAGCGT
3951	GAGCTCTGCT	CTGGCCGAAC	TGGCCACCAA	GACCTTTGGC	AGCAGCGAGA
4001	GCTCTGCCGT	GGACAGCGGA	ACAGCCACCG	CTCTGCCTGA	CCAGGCCAGC
4051	GACGACGGCG	ATAAGGGCAG	CGATGTGGAG	AGCTATAGCA	GCATGCCTCC
4101	CCTGGAAGGC	GAACCTGGCG	ATCCCGATCT	GAGCGATGGC	AGCTGGAGCA
4151	CCGTGAGCGA	AGAGGCCAGC	GAGGACGTGG	TGTGTTGCAG	CATGAGCTAC
4201	ACCTGGACAG	GCGCTCTGAT	CACACCCTGC	GCTGCCGAGG	AGAGCAAGCT
4251	GCCCATCAAC	GCCCTGAGCA	ACAGCCTGCT	GAGGCACCAC	AACATGGTGT
4301	ACGCCACCAC	CAGCAGGTCT	GCCGGACTGA	GGCAGAAGAA	GGTGACCTTC
4351	GACCGCCTGC	AGGTGCTGGA	CGACCACTAC	CGCGATGTGC	TGAAGGAGAT
4401	GAAGGCCAAG	GCCAGCACCG	TGAAGGCCAA	GCTGCTGAGC	GTGGAGGAGG
4451	CCTGCAAGCT	GACCCCCCC	CACAGCGCCA	AGAGCAAGTT	CGGCTACGGC
4501	GCCAAGGACG	TGCGCAACCT	GAGCAGCAAG	GCCGTGAACC	ACATCCACAG
4551	CGTGTGGAAG	GACCTGCTGG	AGGACACCGT	GACCCCCATC	GACACCACCA
4601	TCATGGCCAA	GAACGAGGTG	TTCTGCGTGC	AGCCCGAGAA	GGGCGGCCGC
4651	AAGCCCGCTC	GCCTGATCGT	GTTCCCCGAT	CTGGGCGTGC	GCGTGTGCGA
4701	GAAGATGGCC	CTGTACGACG	TGGTGAGCAC	CCTGCCTCAG	GTGGTGATGG
4751	GCTCAAGCTA	CGGCTTCCAG	TACAGCCCTG	GCCAGCGCGT	GGAGTTCCTG

92/92

4801	GTGAACACCT	GGAAGAGCAA	GAAGAACCCC	ATGGGCTTCA	GCTACGACAC
4851	ACGCTGCTTC	GACAGCACCG	TGACCGAGAA	CGACATCCGC	GTGGAGGAGA
4901	GCATCTACCA	GTGCTGCGAC	CTGGCCCCTG	AGGCCAGGCA	GGCCATCAAG
4951	AGCCTGACCG	AGCGCCTGTA	CATCGGAGGC	CCTCTGACCA	ACAGCAAGGG
5001	ACAGAACTGC	GGATACAGGC	GCTGTAGGGC	CTCTGGCGTG	CTGACCACCA
5051	GCTGTGGCAA	CACCCTGACC	TGCTACCTGA	AGGCCAGCGC	TGCCTGTCGC
5101	GCTGCCAAGC	TGCAGGACTG	CACCATGCTG	GTGAACGCCG	CTGGCCTGGT
5151	GGTGATTTGT	GAAAGCGCTG	GCACCCAGGA	AGATGCTGCC	AGCCTGCGCG
5201	TGTTCACCGA	GGCCATGACC	AGGTACTCTG	CCCCTCCCGG	AGACCCCCCT
5251	CAGCCCGAAT	ACGACCTGGA	GCTGATCACC	AGCTGCTCAA	GCAACGTGAG
5301	CGTGGCTCAC	GACGCCAGCG	GAAAGCGCGT	GTACTACCTG	ACACGCGATC
5351	CCACCACCCC	TCTGGCTCGC	GCTGCCTGGG	AAACCGCTCG	CCATACACCC
5401	GTGAACAGCT	GGCTGGGCAA	CATCATCATG	TACGCCCCTA	CCCTGTGGGC
5451	TCGCATGATC	CTGATGACCC	ACTTCTTCAG	CATCCTGCTG	GCTCAGGAGC
5501	AGCTGGAGAA	GGCCCTGGAC	TGCCAGATTT	ACGGCGCTTG	CTACAGCATC
5551	GAGCCCCTGG	ACCTGCCCCA	AATCATCGAG	CGCCTGCACG	GCCTGTCTGC
5601	CTTCAGCCTG	CACAGCTACA	GCCCTGGCGA	AATTAATCGC	GTGGCCAGCT
5651	GTCTGCGCAA	ACTGGGCGTG	CCTCCTCTGC	GCGTGTGGAG	GCATAGGGCT
5701	AGGAGCGTGA	GGGCTAGGCT	GCTGAGCCAG	GGAGGCAGGG	CCGCTACCTG
5751	TGGAAAGTAC	CTGTTCAACT	GGGCCGTGAA	GACCAAGCTG	AAGCTGACCC
5801	CTATCCCTGC	CGCTAGCCAG	CTGGACCTGA	GCGGATGGTT	CGTGGCTGGC
5851	TACAGCGGAG	GCGACATCTA	CCACAGCCTG	TCTCGCGCTC	CCCCTCGCTG
5901	GTTCATGCTG	TGCCTGCTGC	TGCTGAGCGT	GGGCGTGGGC	ATCTACCTGC
5951	TGCCCAACCG	CTAAA			

FIG. 20D

IN THE PCT RECEIVING OFFICE OF THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s):

Merck & Co., Inc

PCT Serial No.:

To Be Assigned

Case No.: PCT ITR0015Y

US/RO

Filing date:

On Even Date Herewith

Authorized Officer:

For:

HEPATITIS C VIRUS VACCINE

To Be Assigned

Assistant Commissioner of Patents

BOX PCT

Washington, D.C. 20231

NUCLEOTIDE AND/OR AMINO ACID SEQUENCE DISCLOSURE, PCT RULE 5.2

Sir:

As required under PCT Rule 5.2, Applicant respectfully encloses a paper (64 pages) and a computer readable form of the Sequence Listing for the above-identified PCT International Application, filed on even date herewith.

I hereby state that the content of the paper and computer readable forms of the Sequence Listing, submitted in accordance with WIPO and Standard ST.23 and under PCT Rule 13ter.1, respectively, are the same.

Respectfully submitted,

Βv

Sheldon O. Heber Reg. No. 38,179

Attorney for Applicants

Merck & Co., Inc. P.O. Box 2000

Rahway, NJ 07065-0907

(732) 594-1958

SEQUENCE LISTING

<110> Merck & Co. Inc., and Istituto Di Ricerche Di Biologia Molecolare P. Angeletti S.P.A. <120> HEPATITIS C VIRUS VACCINE <130> ITR0015Y <150> 60/363,774 <151> 2002-03-13 <150> 60/328,655 <151> 2001-10-11 <160> 17 <170> FastSEQ for Windows Version 4.0 <210> 1 <211> 1985 <212> PRT <213> Artificial Sequence <220> <223> Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide Met Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly 10 Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly 25 30 20 Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys 40 Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr 60 55 Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp 75 70 Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr 90 85 Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala 110 105 100 Asp Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu 120 125 115 Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu 140 135 130 Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys 155 150 Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met 175 170 Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro 185 190 Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly

200

195

205

```
Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr
                                         220
                      215
Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly
                                    235
               230
Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly
                                250
              245
Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly
                             265
        260
Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile
                                  285
                        280
Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile
                     295
                                300
Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val
                 310
                                    315
Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn
                      330
             325
Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly
          340
                             345
Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe
                         360
                                            365
Cys His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly
                      375
                                        380
Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val
                                     395
                  390
Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met
              405
                                 410
Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys
                                                430
                             425
           420
Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu
                                            445
                         440
Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly
                                        460
                     455
Arg Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly
                                    475
                  470
Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr
                                 490
              485
Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val
           500
                             505
Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp
                         520
                                            525
His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp
                                        540
                      535
Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr
                  550
                                    555
Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro
                                570
              565
Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr
                             585
                                               590
          580
Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn
                                            605
                        600
Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met
                  615
                                       620
Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly
                                     635
                  630
```

	Leu			645					650					000	
	Val		660	Ile				665					6/0		
	Glu	675	Leu				680					685			
	Leu 690	Pro				695					700				
705	Lys				710					715					120
Ala	Ala			725					730					133	
	Lys		740					745					150		
	Ser	755					760					765			
	Ala 770					775					780				
725	ı Ile				790					795					800
	Ser			805					810					812	
	gly		820					825					830		
	/ Val	835					840					845			
	Ser 850					855					860				
86	/ Ala				870					875					880
	Gly			885					890					895	
	e Ala		900					905					910		
	c Asp	915					920					925			
	r Gln 930					935					940				
9.4	r Pro				950					955					900
	r Val			965					970					9/5	
	n Leu		980					985		•			990		
	l Trp	995	;				100	0				100	5		
	n Ile 101	0				101	5				102	0			
10	s Thr 25				103	0				103	5				1040
Th	r Thi			104	5				105	50				105	5
Le	u Trp) Arg	y Val 106		Ala	Glu	Glu	Tyr 106	Val	Glu	ı Val	. Thr	107	Val	. Gly

Asp Phe His Tyr Val Thr G	Gly Met Thr Thr A	sp Asn Val Lys Cys Pro
1075	1080	1085
	.095	1100
Leu His Arg Tyr Ala Pro A 1105 1110		eu Leu Arg Giu Giu Vai 115 1120
Thr Phe Gln Val Gly Leu A		
1125	1130	1135
Cys Glu Pro Glu Pro Asp V 1140	Val Ala Val Leu T 1145	nr Ser Met Leu Thr Asp 1150
Pro Ser His Ile Thr Ala G 1155	1160	1165
Ser Pro Pro Ser Leu Ala S 1170 1	Ser Ser Ser Ala S .175	er Gln Leu Ser Ala Pro 1180
Ser Leu Lys Ala Thr Cys T		
1185 1190	1	195 1200
Leu Ile Glu Ala Asn Leu L 1205	1210	1215
. Thr Arg Val Glu Ser Glu A 1220	1225	1230
Pro Leu Arg Ala Glu Glu A 1235	sp Glu Arg Glu V 1240	el Ser Val Pro Ala Glu 1245
Ile Leu Arg Lys Ser Lys L 1250 1	ys Phe Pro Ala A 255	la Met Pro Ile Trp Ala 1260
Arg Pro Asp Tyr Asn Pro P		
1265 1270		275 1280
Tyr Val Pro Pro Val Val H 1285	1290 1290	1295
Pro Pro Ile Pro Pro Pro A		nr Val Val Leu Thr Glu 1310
Ser Ser Val Ser Ser Ala L 1315	eu Ala Glu Leu A 1320	a Thr Lys Thr Phe Gly 1325
Ser Ser Glu Ser Ser Ala V	al Asp Ser Gly Th	nr Ala Thr Ala Leu Pro
	335	1340
Asp Gln Ala Ser Asp Asp G 1345 1350		er Asp Val Glu Ser Tyr 1360
1345 1350 Ser Ser Met Pro Pro Leu G		
1365	1370	1375
Asp Gly Ser Trp Ser Thr V	al Ser Glu Glu A 1385	a Ser Glu Asp Val Val 1390
Cys Cys Ser Met Ser Tyr T		
Cys Cys Ser Met Ser Tyr T 1395 Ala Ala Glu Glu Ser Lys L	hr Trp Thr Gly Al	a Leu Ile Thr Pro Cys 1405
Cys Cys Ser Met Ser Tyr T 1395 Ala Ala Glu Glu Ser Lys L 1410 1 Leu Arg His His Asn Met V	hr Trp Thr Gly Al 1400 eu Pro Ile Asn Al 415 al Tyr Ala Thr Th	a Leu Ile Thr Pro Cys 1405 a Leu Ser Asn Ser Leu 1420 ar Ser Arg Ser Ala Gly
Cys Cys Ser Met Ser Tyr T 1395 Ala Ala Glu Glu Ser Lys Le 1410 1 Leu Arg His His Asn Met V 1425 1430	hr Trp Thr Gly Al 1400 eu Pro Ile Asn Al 415 al Tyr Ala Thr Th	a Leu Ile Thr Pro Cys 1405 a Leu Ser Asn Ser Leu 1420 ar Ser Arg Ser Ala Gly 35
Cys Cys Ser Met Ser Tyr Ti 1395 Ala Ala Glu Glu Ser Lys Lo 1410 1. Leu Arg His His Asn Met Vo 1425 1430 Leu Arg Gln Lys Lys Val Ti 1445	hr Trp Thr Gly Al 1400 eu Pro Ile Asn Al 415 al Tyr Ala Thr Th hr Phe Asp Arg Le 1450	a Leu Ile Thr Pro Cys 1405 a Leu Ser Asn Ser Leu 1420 ar Ser Arg Ser Ala Gly 35 1440 bu Gln Val Leu Asp Asp 1455
Cys Cys Ser Met Ser Tyr Ti 1395 Ala Ala Glu Glu Ser Lys Li 1410 1. Leu Arg His His Asn Met Vi 1425 1430 Leu Arg Gln Lys Lys Val Ti 1445 His Tyr Arg Asp Val Leu Ly 1460	hr Trp Thr Gly Al 1400 eu Pro Ile Asn Al 415 al Tyr Ala Thr Th 14 hr Phe Asp Arg Le 1450 ys Glu Met Lys Al 1465	a Leu Ile Thr Pro Cys 1405 a Leu Ser Asn Ser Leu 1420 ar Ser Arg Ser Ala Gly 35 au Gln Val Leu Asp Asp 1455 a Lys Ala Ser Thr Val 1470
Cys Cys Ser Met Ser Tyr Ti 1395 Ala Ala Glu Glu Ser Lys L 1410 Leu Arg His His Asn Met V 1425 Leu Arg Gln Lys Lys Val Ti 1445 His Tyr Arg Asp Val Leu Ly 1460 Lys Ala Lys Leu Leu Ser V 1475	hr Trp Thr Gly Al 1400 eu Pro Ile Asn Al 415 al Tyr Ala Thr Th 14 hr Phe Asp Arg Le 1450 ys Glu Met Lys Al 1465 al Glu Glu Ala Cy 1480	a Leu Ile Thr Pro Cys 1405 a Leu Ser Asn Ser Leu 1420 ar Ser Arg Ser Ala Gly 35 au Gln Val Leu Asp Asp 1455 a Lys Ala Ser Thr Val 1470 s Lys Leu Thr Pro Pro 1485
Cys Cys Ser Met Ser Tyr Ti 1395 Ala Ala Glu Glu Ser Lys Le 1410 Leu Arg His His Asn Met Vi 1425 Leu Arg Gln Lys Lys Val Ti 1445 His Tyr Arg Asp Val Leu Leu 1460 Lys Ala Lys Leu Leu Ser Vi 1475 His Ser Ala Lys Ser Lys Pl	hr Trp Thr Gly Al 1400 eu Pro Ile Asn Al 415 al Tyr Ala Thr Th 14 hr Phe Asp Arg Le 1450 ys Glu Met Lys Al 1465 al Glu Glu Ala Cy 1480	a Leu Ile Thr Pro Cys 1405 a Leu Ser Asn Ser Leu 1420 ar Ser Arg Ser Ala Gly 35 au Gln Val Leu Asp Asp 1455 a Lys Ala Ser Thr Val 1470 s Lys Leu Thr Pro Pro 1485

Leu Ser Ser Lys Ala Va 1505 15	10	1515	1520
Leu Glu Asp Thr Val Th	r Pro Ile Asp 1	Thr Thr Ile	Met Ala Lys Asn 1535
Glu Val Phe Cys Val Gl 1540	n Pro Glu Lys (1545	Gly Gly Arg	Lys Pro Ala Arg 1550
Leu Ile Val Phe Pro As 1555	p Leu Gly Val 1	Arg Val Cys	Glu Lys Met Ala 1565
Leu Tyr Asp Val Val Se	er Thr Leu Pro	In Val Val 1580	Met Gly Ser Ser
1570 Tyr Gly Phe Gln Tyr Se	1575 er Pro Gly Gln	Arg Val Glu	Phe Leu Val Asn
1505 15	390	1595	1900
Thr Trp Lys Ser Lys Ly		1610	1012
Cys Phe Asp Ser Thr Va	1625		1030
Ile Tyr Gln Cys Cys As 1635	1640		1645
Ser Leu Thr Glu Arg Le	1655	1660)
Gly Gln Asn Cys Gly Ty	r Arg Arg Cys	Arg Ala Ser	Gly Val Leu Thr
1665 16	570	1675	1000
Thr Ser Cys Gly Asn Th	or Leu Thr Cys	Tyr Leu Lys	Ala Ser Ala Ala
1685		1690	1093
Cys Arg Ala Ala Lys Le	1705		1/10
Gly Leu Val Val Ile Cy	1720		1/25
Ser Leu Arg Val Phe Th	nr Glu Ala Met 1735	Thr Arg Tyr 174	Ser Ala Pro Pro
1730 Gly Asp Pro Pro Gln Pr	ro Glu Tyr Asp	Leu Glu Leu 1755	Ile Thr Ser Cys 1760
1745	750		
Ser Ser Asn Val Ser Val 1765		1770	1//5
Tyr Leu Thr Arg Asp Pr 1780	1785		1190
Thr Ala Arg His Thr Pr	ro Val Asn Ser	Trp Leu Gly	Asn IIe IIe Met
1795	1800	-1 1/	1805
Tyr Ala Pro Thr Leu T	rp Ala Arg Met	Tie Leu Met	THI HIS PHE FHE
1810	1815	182	
Ser Ile Leu Leu Ala G		Giu Lys Ala	Leu Asp Cys Gin 1840
1825	830	1835	
Ile Tyr Gly Ala Cys Ty 1845		1820	1033
Ile Glu Arg Leu His G 1860	1865		18/0
Pro Gly Glu Ile Asn A 1875	1880		1882
Pro Pro Leu Arg Val T	1895	190	U
1000			Tree Tree Leu Dhe
Leu Leu Ser Cln Glv G	ly Arg Ala Ala	Thr Cys Gly	DAS TAT DEG LUC
Leu Leu Ser Gln Gly G	910	1915	1320
Leu Leu Ser Gln Gly G 1905 1 Asn Trp Ala Val Lys T 1925	910	1915	1320

```
Ser Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly
                               1945
           1940
Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu
               1960
                                     1965
       1955
Cys Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn
                1975
                                           1980
   1970
Arg
1985
<210> 2
<211> 5965
<212> DNA
<213> Artificial Sequence
<220>
<223> Non-optimized cDNA sequence encoding SEQ. ID. NO.
<400> 2
gccaccatgg cgcccatcac ggcctactcc caacagacgc ggggcctact tggttgcatc
                                                                      60
                                                                     120
atcactagcc ttacaggccg ggacaagaac caggtcgagg gagaggttca ggtggtttcc
accgcaacac aatccttcct ggcgacctgc gtcaacggcg tgtgttggac cgtttaccat
                                                                     180
ggtgctggct caaagacctt agccggccca aaggggccaa tcacccagat gtacactaat
                                                                     240
gtggaccagg acctcgtcgg ctggcaggcg cccccgggg cgcgttcctt gacaccatgc
                                                                     300
acctgtggca gctcagacct ttacttggtc acgagacatg ctgacgtcat tccggtgcgc
                                                                     360
cggcggggcg acagtagggg gagcctgctc tcccccaggc ctgtctccta cttgaagggc
                                                                     420
tettegggtg gtccactget etgecetteg gggcaegetg tgggcatett eegggetgee
                                                                     480
gtatgcaccc ggggggttgc gaaggcggtg gactttgtgc ccgtagagtc catggaaact
                                                                     540
actatgcggt ctccggtctt cacggacaac tcatccccc cggccgtacc gcagtcattt
                                                                     600
caagtggccc acctacacgc tcccactggc agcggcaaga gtactaaagt gccggctgca
                                                                     660
tatgcagccc aagggtacaa ggtgctcgtc ctcaatccgt ccgttgccgc taccttaggg
                                                                     720
tttggggcgt atatgtctaa ggcacacggt attgacccca acatcagaac tggggtaagg
                                                                     780
accattacca caggogococ ogtoacatac totacctatg gcaagtttot tgccgatggt
                                                                     840
ggttgctctg ggggcgctta tgacatcata atatgtgatg agtgccattc aactgactcg
                                                                     900
actacaatct tgggcatcgg cacagtcctg gaccaagcgg agacggcttgg agcgcggctt
                                                                     960
gtcgtgctcg ccaccgctac gcctccggga tcggtcaccg tgccacaccc aaacatcgag
                                                                    1020
gaggtggccc tgtctaatac tggagagatc cccttctatg gcaaagccat ccccattgaa
                                                                    1080
gccatcaggg ggggaaggca tctcattttc tgtcattcca agaagaagtg cgacgagctc
                                                                    1140
gccgcaaagc tgtcaggcct cggaatcaac gctgtggcgt attaccgggg gctcgatgtg
tccgtcatac caactatcgg agacgtcgtt gtcgtggcaa cagacgctct gatgacgggc
tatacgggcg actttgactc agtgatcgac tgtaacacat gtgtcaccca gacagtcgac
                                                                    1320
ttcagcttgg atcccacctt caccattgag acgacgaccg tgcctcaaga cgcagtgtcg
                                                                    1380
                                                                    1440
cqctcgcagc ggcggggtag gactggcagg ggtaggagag gcatctacag gtttgtgact
ccgggagaac ggccctcggg catgttcgat tcctcggtcc tgtgtgagtg ctatgacgcg
                                                                    1500
ggctgtgctt ggtacgagct caccccgcc gagacctcgg ttaggttgcg ggcctacctg
                                                                    1560
aacacaccag ggttgcccgt ttgccaggac cacctggagt tctgggagag tgtcttcaca
                                                                    1620
ggcctcaccc acatagatgc acacttcttg tcccagacca agcaggcagg agacaacttc
                                                                    1680
ccctacctgg tagcatacca agccacggtg tgcgccaggg ctcaggcccc acctccatca
                                                                    1740
tqqqatcaaa tgtggaagtg tctcatacgg ctgaaaccta cgctgcacgg gccaacaccc
                                                                    1800
ttgctgtaca ggctgggagc cgtccaaaat gaggtcaccc tcacccaccc cataaccaaa
                                                                    1860
                                                                    1920
tacatcatgg catgcatgtc ggctgacctg gaggtcgtca ctagcacctg ggtgctggtg
                                                                    1980
ggcggagtcc ttgcagctct ggccgcgtat tgcctgacaa caggcagtgt ggtcattgtg
ggtaggatta tcttgtccgg gaggccggct attgttcccg acagggagtt tctctaccag
                                                                    2040
                                                                    2100
gagttcgatg aaatggaaga gtgcgcctcg cacctccctt acatcgagca gggaatgcag
                                                                    2160
ctcgccgagc aattcaagca gaaagcgctc gggttactgc aaacagccac caaacaagcg
```

gaggetgetg	ctcccgtggt	ggagtccaag	tggcgagccc	ttgagacatt	ctgggcgaag	2220
cacatgtgga	atttcatcag	cgggatacag	tacttagcag	gcttatccac	tetgeetggg	2280
aaccccccaa	tagcatcatt	gatggcattc	acagcctcta	tcaccagccc	geteaceace	2340
caaagtaccc	tcctgtttaa	catcttgggg	gggtgggtgg	ctgcccaact	cgccccccc	2400
agegeegett	caactttcat	gggcgccggc	atcgccggtg	cggctgttgg	cagcatagge	2460
cttaggaagg	tacttataga	cattctggcg	ggttatggag	caggagtggc	eggegegete	2520
gtggccttca	aggtcatgag	cggcgagatg	ccctccaccg	aggacctggt	caatctactt	2580
cctgccatcc	teteteetgg	cgccctggtc	gtcggggtcg	tgtgtgcagc	aatactgcgt	2640
caacacataa	atccaggaga	gggggctgtg	cagtggatga	accggctgat	agegilegee	2700
tegeggggta	atcatgtttc	ccccacgcac	tatgtgcctg	agagcgacgc	cgcagcgcgt	2760
gttactcaga	tcctctccag	ccttaccatc	actcagctgc	tgaaaaggct	ccaccaglyg	2820
attaatgaag	actgctccac	accgtgttcc	ggctcgtggc	taagggatgt	ttgggactgg	2880
atatocacoo	tattaactaa	cttcaagacc	tggctccagt	ccaagctcct	gccgcagcta	2940
ccaggaatcc	cttttttctc	gtgccaacgc	gggtacaagg	gagtctggcg	gggagacggc	3000
atcatgcaaa	ccacctqccc	atgtggagca	cagatcaccg	gacatgtcaa	aaacggttec	3060
atgaggatcg	tcgggcctaa	gacctgcagc	aacacgtggc	atggaacatt	ccccatcaac	3120
gcatacacca	caaacccctg	cacaccctct	ccagcgccaa	actattctag	ggcgctgtgg	3180
canataacca	ctgaggagta	cgtggaggtc	acgcgggtgg	gggatttcca	ctacgtgacg	3240
ggcatgacca	ctgacaacgt	aaagtgccca	tgccaggttc	cggctcctga	attetteacg	3300
gaggtggacg	gagtgcggtt	gcacaggtac	gctccggcgt	gcaggcctct	cctacgggag	3360
gaggttacat	tccaggtcgg	gctcaaccaa	tacctggttg	ggtcacagct	accatgcgag	3420
cccaaaccaa	atgtagcagt	gctcacttcc	atgctcaccg	acccctccca	catcacagca	3480
gaaacggcta	agcgtaggtt	ggccaggggg	tctccccct	ccttggccag	ctcttcagct	3540
accepttot	ctacaccttc	cttgaaggcg	acatgcacta	cccaccatgt	ctctccggac	3600
actaacetea	tcgaggccaa	cctcctgtgg	cggcaggaga	tgggcgggaa	cateaceege	3660
geegaeeeaa	agaacaaggt	ggtagtcctg	gactctttcg	acccgcttcg	agcggaggag	3720
geggageogg	aagtatccgt	tccqqcqqag	atcctgcgga	aatccaagaa	gttccccgca	3780
gargagagag	tctagacaca	cccggattac	aaccctccac	tgttagagtc	ctggaaggac	3840
ccaactaca	tecetecaat	ggtgcacggg	tgcccgttgc	cacctatcaa	ggcccctcca	3900
ataccacctc	cacggagaaa	gaggacggtt	gtcctaacag	agtcctccgt	gtcttctgcc	3960
ttagcggagc	tegetactaa	gaccttcggc	agctccgaat	catcggccgt	egacagegge	4020
scaacasca	cccttcctga	ccaggcctcc	gacgacggtg	acaaaggatc	cgacgttgag	4080
togtactect	ccatgccccc	ccttgagggg	gaaccggggg	accccgatct	cagtgacggg	4140
tettaateta	ccgtgagcga	ggaagctagt	gaggatgtcg	tctgctgctc	aatgtcctac	4200
acatogaceo	gcgccttgat	cacaccatac	gctgcggagg	aaagcaagct	gcccatcaac	4260
acatggacag	actetttact	gcgccaccat	aacatggttt	atgccacaac	atctcgcagc	4320
gegeegagea	ggcagaagaa	ggtcaccttt	gacagactgc	aagtcctgga	cgaccactac	4380
gcaggcctgc	tcaaggagat	gaaggcgaag	gcgtccacag	ttaaggctaa	actcctatcc	4440
cyggacycyc	cctgcaaget	gacgcccca	catteggeea	aatccaagtt	tggctatggg	4500
gcagaggaag	teeggaacet	atccagcaag	gccgttaacc	acatccactc	cgtgtggaag	4560
geadaggaeg	aagacactgt	gacaccaatt	gacaccacca	tcatggcaaa	aaatgaggtt	4620
ttctatatcc	aaccagagaa	aggaggcgt	aagccagccc	gccttatcgt	attcccagat	4680
ctagagatco	gratatacaa	gaagatggcc	ctctatgatg	tggtctccac	ccttcctcag	4740
atcatastaa	gctcctcata	cogattccag	tactctcctg	ggcagcgagt	cgagttcctg	4800
gregregategg	ggaaatcaaa	gaaaaacccc	atgggctttt	catatgacac	tcgctgtttc	4860
grgaataccc	tcaccgagaa	cgacatccgt	gttgaggagt	caatttacca	atgttgtgac	4920
gactcaacgg	aadccadaca	ggccataaaa	tcgctcacag	agcggcttta	tatcgggggt	4980
tetesets	attcaaaagg	gcagaactgc	ggttatcgcc	gataccacac	gagcggcgtg	5040
-tanganta	actacaataa	caccctcaca	tattacttaa	aggeetetge	agcctgtcga	5100
actacasaaa	tccaccacta	cacgatgete	gtgaacgccg	ccgaccttat	cgttatctgt	5160
getgegaage	reacceases	adacacaaca	agcctacgag	tcttcacgga	ggctatgact	5220
gaaagcycgg	CCCCCCCCC	adacccaccc	caaccagaat	acgacttqqa	gctgataaca	5280
aggracticty	ccaatatata	ggaccegce	gatgcatcag	gcaaaagggt	gtactacctc	5340
calguicat	ccaccaccac	cctcacacac	actacataaa	aaacagctag	acacactcca	5400
accegegate	ancheance	cattatcato	r tatgcgccca	ctttataaa	aaggatgatt	5460
gttaactcct	gyctagycaa	Jacobse	,			

```
ctgatgactc acttcttctc catccttcta gcacaggagc aacttgaaaa agccctggac
                                                                    5520
                                                                    5580
tgccagatct acggggcctg ttactccatt gagccacttg acctacctca gatcattgaa
                                                                    5640
cgactccatg qccttagcgc attttcactc catagttact ctccaggtga gatcaatagg
                                                                    5700
gtggcttcat gcctcaggaa acttggggta ccacccttgc gagtctggag acatcgggcc
                                                                    5760
aggagggaggg ccgccacttg tggcaagtac
                                                                    5820
ctcttcaact gggcagtgaa gaccaaactc aaactcactc caatcccggc tgcgtcccag
                                                                    5880
ctggacttgt ccggctggtt cgttgctggt tacagcgggg gagacatata tcacagcctg
                                                                    5940
tetegtgeec gacceegetg gtteatgetg tgeetactee taetttetgt aggggtagge
                                                                    5965
atctacctgc tccccaaccg ataaa
<210> 3
<211> 5965
<212> DNA
<213> Artificial Sequence
<223> Optimized cDNA encoding SEQ ID NO: 1
<400> 3
                                                                      60
qccaccatgg cccccatcac cgcctacagc cagcagaccc gcggcctgct gggctgcatc
                                                                     120
atcaccagee tgaceggeeg egacaagaae caggtggagg gegaggtgea ggtggtgage
                                                                     180
accgccaccc agagettect ggccacctgc gtgaacggcg tgtgctggac cgtgtaccac
                                                                     240
ggcgccggca gcaagaccct ggccggcccc aagggcccca tcacccagat gtacaccaac
gtggaccagg acctggtggg ctggcaggcc cccccggcg cccgcagcct gacccctgc
                                                                     300
                                                                     360
acctgcggca gcagcgacct gtacctggtg acccgccacg ccgacgtgat ccccgtgcgc
cgccgcggcg acagccgcgg cagcctgctg agcccccgcc ccgtgagcta cctgaagggc
                                                                     420
                                                                     480
agcagcggcg gcccctgct gtgccccagc ggccacgccg tgggcatctt ccgcgccgcc
gtgtgcaccc gcggcgtggc caaggccgtg gacttcgtgc ccgtggagag catggagacc
                                                                     540
accatgcgca gccccgtgtt caccgacaac agcagccccc ccgccgtgcc ccagagcttc
                                                                     600
caggtggccc acctgcacgc ccccaccggc agcggcaaga gcaccaaggt gcccgccgcc
                                                                     660
                                                                     720
tacgccgccc agggctacaa ggtgctggtg ctgaacccca gcgtggccgc caccctgggc
                                                                     780
tteggegeet acatgageaa ggeecaegge ategaeeeca acateegeae eggegtgege
accatcacca ccggcgcccc cgtgacctac agcacctacg gcaagttcct ggccgacggc
                                                                     840
ggctgcagcg gcggcgccta cgacatcatc atctgcgacg agtgccacag caccgacagc
                                                                     900
accaccatcc tgggcatcgg caccgtgctg gaccaggccg agaccgccgg cgcccgcctg
                                                                     960
                                                                    1020
gtggtgctgg ccaccgccac ccccccggc agcgtgaccg tgccccaccc caacatcgag
                                                                    1080
gaggtggccc tgagcaacac cggcgagatc cccttctacg gcaaggccat ccccatcgag
                                                                    1140
gccatccgcg gcggccgcca cctgatcttc tgccacagca agaagaagtg cgacgagctg
                                                                    1200
gccgccaagc tgagcggcct gggcatcaac gccgtggcct actaccgcgg cctggacgtg
                                                                    1260
agcgtgatcc ccaccatcgg cgacgtggtg gtggtggcca ccgacgccct gatgaccggc
tacaccggcg acttcgacag cgtgatcgac tgcaacacct gcgtgaccca gaccgtggac
                                                                    1320
ttcagcctgg accccacct caccatcgag accaccaccg tgccccagga cgccgtgagc
                                                                    1380
cgcagccagc gccgcggccg caccggccgc ggccgccgcg gcatctaccg cttcgtgacc
                                                                    1440
cccggcgagc gccccagcgg catgttcgac agcagcgtgc tgtgcgagtg ctacgacgcc
                                                                    1500
ggctgcgcct ggtacgagct gaccccgcc gagaccagcg tgcgcctgcg cgcctacctg
                                                                    1560
aacacccccg gcctgcccgt gtgccaggac cacctggagt tctgggagag cgtgttcacc
                                                                    1620
ggcctgaccc acatcgacgc ccacttcctg agccagacca agcaggccgg cgacaacttc
                                                                    1680
                                                                    1740
ccctacctgg tggcctacca ggccaccgtg tgcgcccgcg cccaggcccc ccccccagc
                                                                    1800
tgggaccaga tgtggaagtg cctgatccgc ctgaagccca ccctgcacgg ccccaccccc
ctgctgtacc gcctgggcgc cgtgcagaac gaggtgaccc tgacccaccc catcaccaag
                                                                    1860
tacatcatgg cctgcatgag cgccgacctg gaggtggtga ccagcacctg ggtgctggtg
                                                                    1920
ggcggcgtgc tggccgcctt ggccgcctac tgcctgacca ccggcagcgt ggtgatcgtg
                                                                    1980
                                                                    2040
ggccgcatca tcctgagcgg ccgccccgcc atcgtgcccg accgcgagtt cctgtaccag
                                                                    2100
qaqttcqacg agatqqagga gtgcgccagc cacctgccct acatcgagca gggcatgcag
                                                                    2160
ctggccgagc agttcaagca gaaggccttg ggcctgctgc agaccgccac caagcaggcc
```

						2222
gaggccgccg	ccccgtggt	ggagagcaag	tggcgcgccc	tggagacctt	ctgggccaag	2220
cacatgtgga	acttcatcag	cggcatccag	tacctggccg	gcctgagcac	cctgcccggc	2280
aaccccqcca	tegecageet	gatggccttc	accgccagca	tcaccagccc	cctgaccacc	2340
cagagcaccc	tgctgttcaa	catcctgggc	ggctgggtgg	ccgcccagct	ggccccccc	2400
agcgccgcca	gcgccttcgt	gggcgccggc	atcgccggcg	ccgccgtggg	cagcatcggc	2460
ctaggcaagg	tgctggtgga	catcctggcc	ggctacggcg	ccggcgtggc	cggcgccctg	2520
gtggccttca	aggtgatgag	cggcgagatg	cccagcaccg	aggacctggt	gaacctgctg	2580
cccaccatcc	tgagccccgg	cgccctggtg	gtgggcgtgg	tgtgcgccgc	catcctgcgc	2640
caccacataa	gccccggcga	gggcgccgtg	cagtggatga	accgcctgat	cgccttcgcc	2700
agccgcggca	accacgtgag	ccccacccac	tacgtgcccg	agagcgacgc	cgccgcccgc	2760
gtgacccaga	tcctgagcag	cctgaccatc	acccagctgc	tgaagcgcct	gcaccagtgg	2820
atcaacgagg	actgcagcac	ccctgcagc	ggcagctggc	tgcgcgacgt	gtgggactgg	2880
atctgcaccg	tgctgaccga	cttcaagacc	tggctgcaga	gcaagctgct	gccccagctg	2940
cccaacatac	ccttcttcag	ctgccagcgc	ggctacaagg	gcgtgtggcg	cggcgacggc	3000
atcatgcaga	ccacctgccc	ctgcggcgcc	cagatcaccg	gccacgtgaa	gaacggcagc	3060
atgcgcatcg	tgggccccaa	gacctgcagc	aacacctggc	acggcacctt	ccccatcaac	3120
gcctacacca	ccaacccctg	caccccagc	cccgccccca	actacagccg	cgccctgtgg	3180
cacataacca	ccgaggagta	cgtggaggtg	acccgcgtgg	gcgacttcca	ctacgtgacc	3240
ggcatgacca	ccgacaacgt	gaagtgcccc	tgccaggtgc	ccgcccccga	gttcttcacc	3300
gaggtggacg	gcgtgcgcct	gcaccgctac	gcccccgcct	gccgccccct	gctgcgcgag	3360
gaggtgacct	tccaggtggg	cctgaaccag	tacctggtgg	gcagccagct	gccctgcgag	3420
cccgagcccg	acgtggccgt	gctgaccagc	atgctgaccg	accccagcca	catcaccgcc	3480
gagaccgcca	agcgccgcct	ggcccgcggc	agccccccca	gcctggccag	cagcagcgcc	3540
agccagctga	gcgcccccag	cctgaaggcc	acctgcacca	cccaccacgt	gagccccgac	3600
accaacctaa	tcgaggccaa	cctgctgtgg	cgccaggaga	tgggcggcaa	catcacccgc	3660
gtagagagca	agaacaaggt	ggtggtgctg	gacagetteg	accccctgcg	cgccgaggag	3720
dacdadcdcd	aggtgagcgt	gcccgccgag	atcctgcgca	agagcaagaa	gttccccgcc	3780
gccatgccca	tctagaccca	ccccgactac	aaccccccc	tgctggagag	ctggaaggac	3840
cccactaca	tacccccat	ggtgcacggc	taccccctgc	ccccatcaa	ggccccccc	3900
atccccccc	cccaccacaa	gcgcaccgtg	gtgctgaccg	agagcagcgt	gagcagcgcc	3960
ctagccgagc	tggccaccaa	gaccttcggc	agcagcgaga	gcagcgccgt	ggacagcggc	4020
accoccacco	ccctgcccga	ccaggccagc	gacgacggcg	acaagggcag	cgacgtggag	4080
acctacacca	gcatgccccc	cctggagggc	gagcccggcg	accccgacct	gagcgacggc	4140
agetacagea	ccatgagcga	ggaggcagc	gaggacgtgg	tgtgctgcag	catgagctac	4200
acctddaccd	acaccctaat	cacccctgc	gccgccgagg	agagcaagct	gcccatcaac	4260
acceggaceg	acageetget	gcgccaccac	aacatggtgt	acgccaccac	cagccgcagc	4320
accaacctac	accadaagaa	ggtgaccttc	gaccgcctgc	aggtgctgga	cgaccactac	4380
geeggeetge	traarrarat	gaaggccaag	accadcacca	tgaaggccaa	gctgctgagc	4440
atagaagaa	cctacaaact	gaccccccc	cacagegeea	agagcaagtt	cggctacggc	4500
gragagaaga	tacacaacct	gaggaggaag	gccgtgaacc	acatccacag	cgtgtggaag	4560
gccaaggacg	aggagageet	gacccccatc	gacaccacca	tcatggccaa	gaacgaggtg	4620
ttetacatac	aggacaccgc	aaacaaccac	aagcccgccc	gcctgatcgt	gttccccgac	4680
ctagagatac	agecegagaa	gagagataacc	ctgtacgacg	tggtgagcac	cctgccccag	4740
ataataataa	gcgcgcgcga	caacttccaa	tacagccccg	gccagcgcgt	ggagttcctg	4800
gtggtgatgg	gcagcagcca	gaggaggccc	atoggettea	gctacgacac	ccgctgcttc	4860
grgaacaccc	taaccaacaa	cgacatccgc	gtggaggaga	gcatctacca	gtgctgcgac	4920
gacagcaccg	aggeegagaa	agccatcaag	agcctgaccg	agcgcctgta	catcogcogc	4980
erggeceeg	aggeeegeea	ccadaactoc	ggctaccgcc	gctgccgcgc	cagcggcgtg	5040
-benegatia	acagcaaggg	caccataacc	tactacctaa	aggccagcgc	cacctaccac	5100
ctgaccacca	tacacasata	caccatacta	atascacca	ccggcctggt	ggtgatctgc	5160
gccgccaagc	racaggactg	accaract	acctacaca	tgttcaccga	ggccatgacc	5220
gagagegeeg	gcacccagga	ggacgccgcc	carccrart	acgacctgga	gctgatcacc	5280
cgctacagcg	ccccccgg	cataggggg	caycccyayc	gcaagcgcgt	gtactaccto	5340
agctgcagca	gcaacgtgag	cgtggcccac	gacyccaycy	adaccacca	ccacaccccc	5400
acccgcgacc	ccaccacccc	cotggeeege	tacuccuca	agaccgcccg	ccacataate	5460
gtgaacagct	ggctgggcaa	catcatcatg	tacgececa	ccctgtgggc	ccycacyacc	2400

```
ctgatgaccc acttcttcag catcctgctg gcccaggagc agctggagaa ggccctggac
                                                                     5520
tgccagatct acggcgcctg ctacagcatc gagcccctgg acctgcccca gatcatcgag
                                                                     5580
                                                                     5640
cgcctgcacg gcctgagcgc cttcagcctg cacagctaca gccccggcga gatcaaccgc
gtggccagct gcctgcgcaa gctgggcgtg cccccctgc gcgtgtggcg ccaccgcgcc
                                                                     5700
                                                                     5760
cgcagcgtgc gcgcccgcct gctgagccag ggcggccgcg ccgccacctg cggcaagtac
                                                                     5820
ctgttcaact gggccgtgaa gaccaagctg aagctgaccc ccatccccgc cgccagccag
                                                                    . 5880
ctggacctga gcggctggtt cgtggccggc tacagcggcg gcgacatcta ccacagcctg
                                                                     5940
agcogoco gooccogoty gttcatgoty tycotyctyc tyctgagogt gygogtyggo
                                                                     5965
atctacctgc tgcccaaccg ctaaa
<210> 4
<211> 37090
<212> DNA
<213> Artificial Sequence
<220>
<223> MRKAd6-NSmut nucleic acid
<400> 4
catcatcaat aatatacctt attttggatt gaagccaata tgataatgag ggggtggagt
                                                                       60
                                                                      120
ttgtgacgtg gcgcggggcg tgggaacggg gcgggtgacg tagtagtgtg gcggaagtgt
                                                                      180
gatgttgcaa gtgtggcgga acacatgtaa gcgacggatg tggcaaaagt gacgtttttg
                                                                      240
gtgtgcgccg gtgtacacag gaagtgacaa ttttcgcgcg gttttaggcg gatgttgtag
                                                                      300
taaatttggg cgtaaccgag taagatttgg ccattttcgc gggaaaactg aataagagga
                                                                      360
agtgaaatct gaataatttt gtgttactca tagcgcgtaa tatttgtcta gggccgcggg
                                                                      420
gactttgacc gtttacgtgg agactcgccc aggtgttttt ctcaggtgtt ttccgcgttc
cgggtcaaag ttggcgtttt attattatag gcggccgcga tccattgcat acgttgtatc
                                                                      480
catatcataa tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt
                                                                      540
gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata
                                                                      600
tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc
                                                                      660
cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc
                                                                      720
attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt
                                                                      780
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt
                                                                      840
atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca
                                                                      900
togotattac catggtgatg oggttttggc agtacatcaa tgggcgtgga tagcggtttg
                                                                      960
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc
                                                                     1020
aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg
                                                                     1080
gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg
                                                                     1140
                                                                     1200
cctggagacg ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc
tccgcggccg ggaacggtgc attggaacgc ggattccccg tgccaagagt gagatctgcc
                                                                     1260
                                                                     1320
accatggcgc ccatcacggc ctactcccaa cagacgcggg gcctacttgg ttgcatcatc
actageetta caggeeggga caagaaccag gtegagggag aggtteaggt ggtttecace
                                                                     1380
                                                                     1440
gcaacacaat ccttcctggc gacctgcgtc aacggcgtgt gttggaccgt ttaccatggt
gctggctcaa agaccttagc cggcccaaag gggccaatca cccagatgta cactaatgtg
                                                                     1500
gaccaggacc tcgtcggctg gcaggcgccc cccggggcgc gttccttgac accatgcacc
                                                                     1560
tgtggcagct cagacettta ettggtcacg agacatgetg acgteattee ggtgcgcegg
                                                                     1620
cggggcgaca gtagggggag cctgctctcc cccaggcctg tctcctactt gaagggctct
                                                                     1680
togggtggtc cactgotetg coettogggg cacgotgtgg gcatcttccg ggctgccgta
                                                                     1740
tgcacceggg gggttgcgaa ggeggtggac tttgtgcccg tagagtccat ggaaactact
                                                                     1800
atgoggtote eggtetteac ggacaactea tecceeegg eegtacegea gteattteaa
                                                                     1860
                                                                     1920
qtggcccacc tacacgctcc cactggcagc ggcaagagta ctaaagtgcc ggctgcatat
gcagcccaag ggtacaaggt gctcgtcctc aatccgtccg ttgccgctac cttagggttt
                                                                     1980
ggggcgtata tgtctaaggc acacggtatt gaccccaaca tcagaactgg ggtaaggacc
                                                                     2040
attaccacag gcgccccgt cacatactct acctatggca agtttcttgc cgatggtggt
                                                                     2100
tgctctgggg gcgcttatga catcataata tgtgatgagt gccattcaac tgactcgact
                                                                     2160
```

acaatcttgg	gcatcggcac	agtcctggac	caagcggaga	cggctggagc	gcggcttgtc	2220
gtgctcgcca	ccgctacgcc	tccgggatcg	gtcaccgtgc	cacacccaaa	catcgaggag	2280
gtggccctgt	ctaatactgg	agagatcccc	ttctatggca	aagccatccc	cattgaagcc	2340
atcagggggg	gaaggcatct	cattttctgt	cattccaaga	agaagtgcga	cgagctcgcc	2400
gcaaagctgt	caggcctcgg	aatcaacgct	gtggcgtatt	accgggggct	cgatgtgtcc	2460
			gtggcaacag			2520
			aacacatgtg			2580
agcttggatc	ccaccttcac	cattgagacg	acgaccgtgc	ctcaagacgc	agtgtcgcgc	2640
tcgcagcggc	ggggtaggac	tggcaggggt	aggagaggca	tctacaggtt	tgtgactccg	2700
ggagaacggc	cctcgggcat	gttcgattcc	tcggtcctgt	gtgagtgcta	tgacgcgggc	2760
tgtgcttggt	acgagctcac	ccccgccgag	acctcggtta	ggttgcgggc	ctacctgaac	2820
			ctggagttct			2880
ctcacccaca	tagatgcaca	cttcttgtcc	cagaccaagc	aggcaggaga	caacttcccc	2940
tacctggtag	cataccaagc	cacggtgtgc	gccagggctc	aggccccacc	tccatcatgg	3000
gatcaaatgt	ggaagtgtct	catacggctg	aaacctacgc	tgcacgggcc	aacacccttg	3060
ctgtacaggc	tgggagccgt	ccaaaatgag	gtcaccctca	cccaccccat	aaccaaatac	3120
atcatggcat	gcatgtcggc	tgacctggag	gtcgtcacta	gcacctgggt	gctggtgggc	3180
ggagtccttg	cagctctggc	cgcgtattgc	ctgacaacag	gcagtgtggt	cattgtgggt	3240
			gttcccgaca			3300
ttcgatgaaa	tggaagagtg	cgcctcgcac	ctcccttaca	tcgagcaggg	aatgcagctc	3360
gccgagcaat	tcaagcagaa	agcgctcggg	ttactgcaaa	cagccaccaa	acaagcggag	3420
gctgctgctc	ccgtggtgga	gtccaagtgg	cgagcccttg	agacattctg	ggcgaagcac	3480
			ttagcaggct			3540
			gcctctatca			3600
agtaccctcc	tgtttaacat	cttggggggg	tgggtggctg	cccaactcgc	ccccccagc	3660
gccgcttcgg	ctttcgtggg	cgccggcatc	gccggtgcgg	ctgttggcag	cataggcctt	3720
			tatggagcag			3780
gccttcaagg	tcatgagcgg	cgagatgccc	tccaccgagg	acctggtcaa	tctacttcct	3840
gccatcctct	ctcctggcgc	cctggtcgtc	ggggtcgtgt	gtgcagcaat	actgcgtcga	3900
			tggatgaacc			3960
			gtgcctgaga			4020
			cagctgctga			4080
			tcgtggctaa			4140
tgcacggtgt	tgactgactt	caagacctgg	ctccagtcca	agctcctgcc	gcagctaccg	4200
ggagtccctt	ttttctcgtg	ccaacgcggg	tacaagggag	tctggcgggg	agacggcatc	4260
atgcaaacca	cctgcccatg	tggagcacag	atcaccggac	atgtcaaaaa	cggttccatg	4320
aggatcgtcg	ggcctaagac	ctgcagcaac	acgtggcatg	gaacattccc	catcaacgca	4380
tacaccacgg	gcccctgcac	acceteteca	gcgccaaact	attctagggc	gctgtggcgg	4440
gtggccgctg	aggagtacgt	ggaggtcacg	cgggtggggg	atttccacta	cgtgacgggc	4500
atgaccactg	acaacgtaaa	gtgcccatgc	caggttccgg	ctcctgaatt	cttcacggag	4560
gtggacggag	tgcggttgca	caggtacgct	ccggcgtgca	ggcctctcct	acgggaggag	4620
gttacattcc	aggtcgggct	caaccaatac	ctggttgggt	cacagctacc	atgcgagccc	4680
gaaccggatg	tagcagtgct	cacttccatg	ctcaccgacc	cctcccacat	cacagcagaa	4740
acggctaagc	gtaggttggc	cagggggtct	ccccctcct	tggccagctc	ttcagctagc	4800
cagttgtctg	cgccttcctt	gaaggcgaca	tgcactaccc	accatgtctc	tccggacgct	4860
gacctcatcg	aggccaacct	cctgtggcgg	caggagatgg	gcgggaacat	cacccgcgtg	4920
			tctttcgacc			4980
gagagggaag	tatccgttcc	ggcggagatc	ctgcggaaat	ccaagaagtt	ccccgcagcg	5040
atgcccatct	gggcgcgccc	ggattacaac	cctccactgt	tagagtcctg	gaaggacccg	5100
gactacgtcc	ctccggtggt	gcacgggtgc	ccgttgccac	ctatcaaggc	cccccaata	5160
			ctaacagagt			5220
			tccgaatcat			5280
gcgaccgccc	ttcctgacca	ggcctccgac	gacggtgaca	aaggatccga	cgttgagtcg	5340
tactcctcca	tgcccccct	tgagggggaa	ccgggggacc	ccgatctcag	tgacgggtct	5400
tggtctaccg	tgagcgagga	agctagtgag	gatgtcgtct	gctgctcaat	gccctacaca	5460

ttgagcaact ctttgtgeg caccataac atggtttat cacacact tcgcagca ggcttgcggc agaagaagag cacctttgac agactgcaa gactgcaagagaggc tccacagta aggctgacac ccctaccgg 540 agagaagaggt gccccaagta aggctaaact cctatcgta gaggaagacct gcaagctgac gccccacat tcggccaaat ccaagtttg gtggaagagc ttgctggaag acactggaa gccccaagta aggcataact cctatcgta gaggaagacct gagacagcc gtaacacact cagcaagcc gttaaccaca tccactccgt gtggaaggac ttgctggaag acactgtgaa accaattgaa accacata tggcaaaaaa tgaggttttc 588 tgtgtggaag acactgtgaa gaggccgtaaa ccaaccata tggcaaaaaa tgaggttttc gtgtgtgaga gatgaggccgtaag cagccccc tattgatgtg tctccaccct tcctcaggtc gtgatgggc tctcatacgg attccagtac tctcctggg agcagtggg gttcctggtg attcagagaa gatgaccct tattgatgtgt tctccaccct tcctcaggt gtgatgggc cccagaagaga accccatag ggcttttca atgacactcg tgttttca tggcccccagaagacaga catccgtgt gaggagtaa tttaccaatag ttgttgacttg gccagaggac accagacagg cataaaatcg ctcacagagc ggctgcggg gcgggggggggg							
ttgagcaact ctttgctgcg caccataac atgstttats cacacactc tcgcagcca 558 goctgcgcca gagaagagg caccttatga cagatcgaas tctctgacaga cactaccgts 570 gaggaagcct gcaagctgac gccccacat tcgccaaat ccaagttag ctatagggac 576 aaggaagcc gcaagctgac gccccacat tcgccaaat ccaagttag ctatagggac 576 aaggaagccc ggaactatc cagcaaggcc gttaaccaca tccactccgt gtggaaggac 576 tcgtggaag caccattgac accacttgac tccactccgt gtggaaggac 576 tcgtggaag cacattggac accaattgac accacatta tggcaaaaa tgaggttttc 588 tggtgccaac cagagaaagg agccgcagc ttatetgtatt ccaagatcg 576 gagatccgtg tatcggaaa gatggcccc tatgatgtg tcccaacct tcctcaaggtc gtgatgaggac accacatca accacatca tggcaaaaaa tgaggttttc 588 tggtgcacac cagagaagga gatggccccc tatgatgtg tctccaacct tcctcaaggtc gtgatgaggac accacatcag gcttttcat atgacaccc tcctcaggc gtgatgag gtccccgaag cacacagga aaccccatt gaggaatcaa tttaccaatt ttgagagtc tccaaaggac gccccacaatt accacagac gctttatat cagggggcc 624 ccgagatca gagcacgcag cacagaagga gcagcaggag cacagaatcg cacagaagga gcgcggggg ctacaggtc tcctgagac tggggtgct 530 ggaagctcc aggatca gagctgcgg accacagt tatcagaag cctctgaaga cggcgggag cgcgggga cgcggggag cgcggggag cgcggggag cgcggggag cgcgggga caggatcac gagcacac cctcacatgt tatcagaag ccttgaagac tgtgcagga accagaagac gcgcacaaga cagacagga cacagaagac cacagaagaa aacccatgt aaccagaaga gagatgaca gagcaggaagac cacagaagac gcgcggag ccacagatc tccacagaagaga cagaacagac gcgcgagaa cagagaagac gcggagagac cacagaagac gagcaggaac cacagaagaa gagcacaggagaac cacagaagaaga gcgagaagac cacagaagaaga aacccaagaaga gagaactagaa aacccaagaaga aacccaagaaga aacccaagaagaaga aacccaagaa aacccaagaagaagaaga cacagaagaaga cacacagaagaaga cacagaagaaga cacaagagaagaa cacagaagaagaagaa aacccaagaagaagaagaa aacccaagaagaagaagaaaagaagaagaagaagaagaag	tggacaggcg	ccttgatcac	gccatgcgct	gcggaggaaa	gcaagctgcc	catcaacgcg	5520
ggectgecge agaagaaggt eacetttgac agactgecag tectggacga ceattacegg 54gaaggagget geaaggteca aggaaggagg geceaagta aggecaaget ceatacegt 57gaaggaagcte geaagtgac gececaaatt aggecaaat ceaagttgg catagggaggac 562ttgetggaag acactgtgac accaactgac gtaaccaca tecacteegt gtggaaggac 562ttgetggaag acactgtgac accaattgac accaccacat tggcaaaaaa tgaggtttt 588 ggatgecgt tatgggaag aggecgtaag cagecgec tategtatt cecagaattg ggagtacgtc gtatggaag agaggecet cataggtgt tecacacet tectcaggt 600 gtgatgggt tatggagaag atggecget cettetggg agaggaggac 600 gtgatgggac accaacagg accaccatg ggetttetaa taggaaatacac cggagaagga accaacaga gaccaccatg ggetttetaa taggacatga gttectggg 612 etaacaggat accaagaga accaccatg ggetttetaa taggacatga gttectggg 612 etaacaggac ceagacaggc cataaaateg tecacagaga ggecgegaag ceagacaggc cataaaateg tecacagaga ggecgegaga cegagaggac cataaaateg tecacagagag ggecgegaga cggegggaga cacacagtg gggtaaacac cetacaatgt tactggagg ceceggag cggegggagac cagaaggac cacacacagt geggagaca cacacacag gactgggagac tacagagac ggcgeggaga cecacagaga gacggggagac cacacagaga agcggcagag ctacagagac accaccagaga agcggagac cacacagaga cacacaca	ttgagcaact	ctttgctgcg	ccaccataac	atggtttatg	ccacaacatc	tcgcagcgca	5580
gaegtgecte aggagatgaa gegeaaggeg tecacagtta aggetaaate cetateceta aaggagagtee gaaactate cageaagge gttaaccace tecacteegt gtgagaaggac 528 aaggacgtee ggaacetate cageaaggee gttaaccace tecacteegt gtgagaaggac 528 ttectgagaa cacettgaa accaattgac accactcate tggeaaaaaa tgaggtttee 588 gtgtecacae cagagaagga gaggecgtaag caagecegee ttategtatt eccagatety gtgagaagaag aggecgtaag caagecegee ttategtatt eccagagteg gtgagaaggac 628 atacetagga atacaagga aggecete tatgatgtyg tetecacecet tecteaggte gtgagaaggac cetacatagga atacaagga accaagtag ggetttetat atgacacteg steeteggga accaaggacga caccagtag gggaggateaa tttaccaagg ttecteggg accaccaggag cagaaagac caccattg tactgaggg ggetttatat eggaggtee 624 ecgagacga caccagte tategeggg ggetttatat eggaggtee 624 ecgagageta gggagagaca cagacagge caccagtt tategeggg ggetttatat eggagggtee 624 ecgagageta gggagagaca eggagagaca eaccagtgt tategeggg ggetttatat eggaggetee 624 ecgagageta gggagagaca eccagagaga eccagagaga eccagagac eccagaga eccagaga eccagagaga eccagagaga eccagagag eccacagat eacagaga eccagagaga eccagagaga eccacagaga eccacagat tategagget gggagagac ettytogag 648 ectetyeec eccegggga eccecacaga eccagate eccagagat eccagagate ettytogaga 648 ectetyeec ecceggga eccaccaga gaegagaga eccagagate eccacagaga eccagagate eccacagate eccagagaga eccagagaga eccacagate gaeagaga eccagagate ettetetaca eccaccate eccaccate eccaccate eccaccagate eccacagate eccagagate eccapagate eccapagate eccapagaga eccapagate eccapagate eccapagate eccapagate eccapagate eccapagagate ettetetacae eccaccate eccacagate eccapagea eccapagate eccapagaga eccapagaacate tecagagate eccapagagaa eccategagaga eccapagaacate tecagagaga eccapagagaacate eccapagagagaace eccapagaga eccapagagaga eccapagaga eccapagagaga eccapagaga eccapagagaga eccapagagagaga eccapagagagagagagagagagagagagagagagagagag	ggcctgcggc	agaagaaggt	cacctttgac	agactgcaag	tcctggacga	ccactaccgg	5640
agagaagect geaagetgae gececeaeat teggeaaat ceaagettgg chatgggga 592 ttgetggaag acactgtgae accaattgae accacatca tggeaaaaaa tggaggttte 588 ttgetgeaag acactgtgae accaattgae accacatca tggeaaaaaa tggaggttte 588 ggagtecgt tatggagaa gatggeecte atatggtgg tecteaeccet tecteagge 600 ggagtecgt tatggagaa gatggeecte tatggtgg tecteaeccet tecteagge 600 gtaptagget ceteataegg attecagtae teteetggg ageagetga gtteetggg 612 staacggtae accaaaagaa aaacceaatg ggetttetaa tagacacteg etgetttegae teaeacgag accacagaga catecagtg gagagatea tttaccaatg ttgtgactig 618 geeceegaag cegaacagg cataaaateg etaeaagag ggetttatat taggggteet 62 cegaacagge cataaaateg etaeaagag ggeetgag gggeggege 618 gegaagete cagacagge gaatagaetg etaeaagag ggettytegt tatetydgag 618 gegaagete cagacagge cataaaateg etaeagaga ggeetgggg cggeggggg 62 gegaagete cagacagge cataaaateg etaeagaga ggeetgggg 618 gegaagete cagacagge cataaaateg etaeagaga ggeetggggg 62 gegaagete cagacagga cacacagg caccacaga categagge categagga cecaaggaga cecaaggag caccacaga categagge categagge categagge categagge categagge 62 gegaagete cagacagge cecacaga gatagaacg categagge tategtggaaa caccacagga accacagga acgeggegga cecacagaa gagatgaga cacacagga acgeggggga cecaaggaga cacacagga acgeggggg 64 actettgee cecagggg cegacaga geggggaga cacacagga acgeggagga cacacacaga actegagaga cacacacaga actegagaga cacacacaga acacacaga acacacac	gacgtgctca	aggagatgaa	ggcgaaggcg	tccacagtta	aggctaaact	cctatccgta	5700
aaggacgtce ggaacetate cageaaggee gttaaceace tecaetcegt gtggaaggac 552 ttgetggaag acactgtgaa accaattga accacata tgagatatte 588 tgtgtecaac cagagaaagg aggecgtaag cagecgee ttategtatt eccagateg 594 ggagteegtg tategagaa gatggecete tatgatgtg teceaecet tecteaggte 666 atacetgga aateaagga aggeegtaag cegeegeegge ttategatt eccagategg 666 aatacetgga aateaaagaa aaacecatg ggettttetat atgacacteg etgettegge 662 actacetgga aateaaagaa aaacecatg ggettttetat atgacacteg etgettegge 624 cegagatea cegagaaegg cataaaateg etcacagagg ggetttatat egggggteet 624 cegagateat caaaaggca gaactegggt tateggegg ggetttatat eggggggteet 624 cegagateat caaaaggca gaactegggt tateggegg ggetttatat eggggggteet 624 cegagateat gaggtaaaca ceteacatgt tatetgaagg cettetgagg etgetgggg aggaagetee aggactgaa gatgetegtg aacgeegegg geettytegt tateetggaa geegaggate cecaagagga geeggegge etaegagtet tecaggage tateggage 648 actetegee cecaaggagg eegeegegg etaegagtet tecaggage tatectggg 648 actetegee cecaaggagg eegeeggeg etaegagtet tecaggage tatectggg 648 actetegee cecaaggage eegeeggeg etaegagete tecaggagg actetegage 648 actetegee cecaaggag eegeeggeg etaegagete tecaggagg actetege 654 dgtteetea atgtgteggt egeeacaga geateagage actetegagt tecatgage 648 actetegge taggeacat tateatgat gegeeacat tggagaaaggagaggaggggggggggggggggggg	gaggaagcct	gcaagctgac	gccccacat	tcggccaaat	ccaagtttgg	ctatggggca	5760
ttgtctagaag acactgtgac acaattgac acaactata tggcaaaaaa tgaggttttc 584 gtgttccaac cagagaaga gatggcctaag ccagccacgc ttatactgtatt cccaqatctg gtgatgtggc cctcatacgg attccagtac ttcctgggc aggagtcag ttctcagtc 600 gtgatgggct cctcatacgg attccagtac ttcctgggc aggagtcag attccagtac 600 gtgatgggct cctcatacgg attccagtac ttctctggc aggagtcag attccagtac 600 gtgatgggcd aatcacagga attccagtac ttctcaggca ggcttttat tggacatcg ctcaacaggc gctcaaaaatcg ctcacaaggc gcttttat tggagagtcat ttggacacag cccgagaacga catccqtgtt gaggagtcaa tttaccaatg ttgtgacttg ctgacacagacgacgacgacgacgacgacgacgacgacgacg	aaggacgtcc	ggaacctatc	cagcaaggcc	gttaaccaca	tccactccgt	gtggaaggac	5820
gyagtcogtc tatgogaaa gagcogtaag ccagcogcc ttatcquatt cccagatctg gyagtcogtg tatgogaaa gagcogctc tatgagtgg tetcacacct tectcagtc 90 gyagtgogct cctcatacgg attrocagtac tectcoggc agcagtcag gttcctggt 606 aatacctgga aatcaaagaa aaacccaatg ggcttttcat atgacactg ttgttcgac 618 geccccqaag ccagaacga catccqtgt gaggagtcaa ttaccagta ttgtgaactg gccccqaag ccagaacga catcaaatg ctcacagag ggctttatat cgggggtcat 618 gcccccqaag ccagaacgac catcaactg tatcgcoggt gcgcgggggg cgggggggc cagaactgcggt tatcgccggg ggctttatat cgggggtcat 626 gcgaagctcc aggactgcac cctcacatgt tacttgaagg cctctgcagc ctgttcgac 636 gcgaagctcc aggactgcac gatcctgtg aacgccgcgc gccttgtcgt tatctgaag cggaggggaag cccaagaagaa cccaagagag cccagagacgac catcaagatg tactgaggac accagagaga cccagagaga cccaagagaga cccagagaga cccagagaga cccagagaga cccagagaga cccagagaga cccagagaga cccagagaga cccacagaga gatcagaga actuggagac actccaatg fatctctgcc cccccggga ccccccaa ccagaatacg acttggagac gataactaa fgttcggt cgccacagat gcatcaaggac acatcagaga atgatacaa catccagt fatctcatga gagactacac cttctccac cctccagac gggggggggg	ttgctggaag	acactgtgac	accaattgac	accaccatca	tggcaaaaaa	tgaggttttc	5880
ggagtacegty tatgegagaa gatgegeete tatgatgtg tetecaceet teeteagt 600 gatgatggget ceteataegg attecetgg 600 gatateetgga acteaaagga aatecaagga attecatgga ggettttaat atgacaeteg ctgtttega 612 georgaagtega gettetgatg 612 georgaagtega georgaagtega gtteetgga 612 georgaagtega ceteaagge catacaagga ggettttaat egggggteet 624 ctgactaatt caaaaggge gaactgeegt tateegggt gettetgat eggggagaete caggatagae ceteacatgt tatettgaagg cetetgage ctgtgagaet caggagagtee gaggagatega gggagatega gggagatega gggagatega gggagatega gggagatega gggagagaete aggagagaete aggagagaete caggagagae ceteacatgt tatetgaagg cetetgagae ctgtgagaga cecacagagaga cattgagaga gatagategagata cacaccaca atgtgteega cagcecacaa gagagagaa cagttgagaca cactccaagt figticeca acaccecac cecacagagaga cagtgagaaa cagcagagaa cactccagagagaatatacaca catteteegagagaacat tatacagtaa ggaccacatt gtgggaaaa cactcagagagaacat tatgaacaga cactcagagagaacatagaa cactcacagagagaacatagaa cactcacagagagaacatagaaca taggagagaaca cactcagagagaacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacacaagagagaacaagagagaacaagagagaacaac	tgtgtccaac	cagagaaagg	aggccgtaag	ccagcccgcc	ttatcgtatt	cccagatctg	5940
atacatgga aatcaaaga atccagtgat ggetttteat atgacatcg ctgtttgac 612 caacggtca ccgagaaaga catccgtgtt gaggagtcaa tttaccaatg ttgtgacttg cccccgaag ccagacagga catcaaatgt ggetttteat atgacactcg ctgtttgac 618 gcccccgaag ccagacagga catcaaatgt gtacagaga ggettatat cgggggtcat ctgactaatt caaaaggga gaactgcggt tatcgagag ggcttatat cgggggtcat cggactagct ggggtaacac cctcacatgt tatctgaagg cccttgaagc cggcgtgagc gcgggaaacac ccccacatgt tactgaagg cgccttgagg tggcgtgtg 630 acgactagct ggggtaacac cctcacatgt tactgaagg cccttgaagc cggcgtgag acgactgaga cccacagagga cacgcgcaa caagaatag cctctgaagc tatgactagg tatcttgcc cccccgggga cccgcccaa ccagaatacg acttggagc tatacatca ggtgtaccac acccccct cgcacgggc ggtgggaaa cactggaga caagggggga tatgatcaag aggctggt ggcgcacacat tactactga tactgagaa aagggtgaa caactcagg tatgatcaca atggtaggt cgccacaga ggataaaga actggggaag gatgattcg ggtgtaccac cacccccct cgcacgggct ggtgggaaa cagtagaca cactcaagt tatgatcact tattctact cctttagac caggagaaac ttgaaaaaga cctggactg 672 aagatcacgg gggcctgtta ctccattgag cacttgagac tacaggaa ctgggaga ccggccagg caggtacagg ttagacacat ttcactcat agttactcc caggtgaaa caggtagaca ccggccagg agcttcaggc taggctact gtccagggg ggagggcg cactttgag caattgagga gg ggcttcaggc cagggaaac tagcacgagg ggaggggcg cactttgag caataggga gg ggcttgcagg cagggaaac caaccagaa caccacaac accaccacaa caccggaaact cgggccagg ggcttcacgg gctggtcgt tgctggttac agcggggagg acatatatac accggctgc gcccaggtgtg caggtagaa acaacaaa ctcaaccaaca accaggta ggtgccaga gccgccgt cccggcggt aatgaggaga cacaccacagg ggaaggaca cacaccaaca accaccacaa accaccacaa agtggggaga acatatatca cggcggggaggaggaggaggaggggggaggggggggg	ggagtccgtg	tatgcgagaa	gatggccctc	tatgatgtgg	tctccaccct	tcctcaggtc	6000
aatacatgga aatcaaagaa aaaccccatg ggcttttcat atgacactcg ctgtttcgac tcaaaggtca ccgagaaga catccgtgtt gagagatcaa tttaacaatg ttgtgaacttg fagaccccgaagacgac cataaaaatcg ctaaaaagac ggctttaata cgggggtcct ctgataatat caaaagggca gaactgcggt tatccgcggt gccgcgcgag cggcgtgctg gcgagagctcc aggactgca gaactgcggt aactgcaggt gccttgcagc ctgtcaggc ggagagccc aggactgcac gatgctcgtg aacgccgggag ctctcgagc ctgtcgagc ctgtcgagc gcgagagccca aggactgcac gatgctcgtg aacgccgggagc ctcacgagtct tcacgaggcg tatacatag gcgagagcac cccaagaataga cccagaaatag ctctggagc tatacataga gagagagcac caccaccacct cccagagagc cacgagagac acagaataga actggagac atacataga fatgatcaca cacccccct cgcacgag gatcaaggaa aaagggtgta ctacctacac gagactacac cacccccct cgcacgggc ggatgagaa caggtagaac cactcagtc aactccaggc tagactaag ggactaact atcatgtat gcgccacatt tgtgggaag gatgattctg for atacatagac actccagt ttacatagac aggactacac caccagat ctcatagac cacctggac ttagacaac actccagt cagaactacac ggactacac ctcatagac cacctggac ttagacaacat ttcaatgaa cacatgaacac cacctcacc cagaatacac ggactgacac cacctagac ggactgacac cacctagac cacctagac ggactgacac ttcaactagac cacctagac ttcaactagac gactgacaca cacctagac cacctagac cacctagacc cacctagacc cacctagaccc cacctagaccc cacctagaccc cacctagaccc cacctagaccc cacctagaccc cacctagaccc cacctagaccc cacctagacccc cacctagacccccccccc	gtgatgggct	cctcatacgg	attccagtac	tctcctgggc	agcgagtcga	gttcctggtg	6060
ccaccgaca ccagaacaga catecgtgtt gaggaqtcaa tttaccaatg ttgtgacttg gccccgaag ccagacaggc cataaaatcg ctcacagaga ggctttatat cgagaggcc 624 ctgactaatt caaaagggca gaactgcggt tactgcaggt gccgcgcgag cggcgtgtg gcgaagctcc aggactgcac gatectgtg tacttgaagg cctctgcaagc ctgtcgagct gcgaagctcc aggactgcac gatectgtg aacgcgcgg gccttgtgt tacttgtgaa gcgaggggaa cccaagagga cgcggcgag ctacgagtct tcacggaggc gctttgtg tacttgtgaa tactctgcc ccccgggga ccccaccac ccagaatacg gctttgtgt tacttgtgaa tactctgcc cccccgggga cccacacac caagaatacg acttggaagct gataacatca fgttcctcca attgtteggt cgcccacaga gaataagca aaagggtgta ctacctaac cgtgatccca ccaccccct cgcacgggt gcgtgggaaa aaagggtgta ctacctaac cgtgatcaca caaccccct cgcacgggt gcgtgggaaa cagcagaacacacacacacacacacacacacac	aatacctgga	aatcaaagaa	aaaccccatg	ggcttttcat	atgacactcg	ctgtttcgac	6120
cccccqaag ccagacaggc cataaaatcg ctcacagagc ggctttatat cgggggtcct 624 ctgactaatt caaaagggca gaactgcggt tactcgagg cccgcgag cggggtgctg 636 acgactagct gcggtaacac cctcacatgt tactcgagg cccctgagg ctgtcgagct 636 gcgaagctcc aggactgcac gatgctcqtg aacgccgcg gccttgtcgt tactcgtgaa dccaaggag cccaagagga cccagagga cccggcgaga cccagagga cccagagga cccagagga cccagagga cccagagga cccagagga cccagagga cccagaggaga cccagaggaga cccagaggaga cccagaggaga cccagaggaga cccagaggaga cccagaggaga cccagaggaga cccagaggaga cccagagagag	tcaacggtca	ccgagaacga	catccgtgtt	gaggagtcaa	tttaccaatg	ttgtgacttg	6180
acgactagct geggtaacac ceteacatgt tacttgaagg cetetgage ctytegage (54) cagagactoc aggactgcac gatgetegtg aacgecgeeg geettgtegt tatetgtgaa (64) cagagageeggaa cecaagagag cecageegga cetacgagtet taceaggage (64) caccegggga cecaagagee cacagataca cateaggagee tatgateegga cecageggeegga cecacagataca cateaggageegggaacaca cateagageegggaacacacacacacacacacacacacacaca	gcccccgaag	ccagacaggc	cataaaatcg	ctcacagagc	ggctttatat	cgggggtcct	6240
acgactagct geggtaacac ceteacatgt tacttgaagg cetetgage ctytegage (54) cagagactoc aggactgcac gatgetegtg aacgecgeeg geettgtegt tatetgtgaa (64) cagagageeggaa cecaagagag cecageegga cetacgagtet taceaggage (64) caccegggga cecaagagee cacagataca cateaggagee tatgateegga cecageggeegga cecacagataca cateaggageegggaacaca cateagageegggaacacacacacacacacacacacacacaca	ctgactaatt	caaaagggca	gaactgcggt	tatcgccggt	gccgcgcgag	cggcgtgctg	6300
aggaagatccaggactgaagcaggacgagacatetytaagfatetytaaaggacgggaagcccaagaggacctacgaggacactetytaaggfatetytaaggtattetytecccccaccaggagcccaccaacacaagagtcacttygagcgataacatcacgtyatccaccaccccctcgccacagatgcatgagaaaaaggytgtactacctcaagtcgtyatccaccaccccctcgccacagatgcatgagaaacagctagaaacactccagt666cagtactaacttcttetccatccttectagacaggagaaatytgagcaaggatyattetg672cagattaacggggcetytactcattyagcacttyacccactcagatcatyaagaccatyaagacg684ctcataggcttatgagcaactttatcatyattggagaaacttyaaaaagacacttygagt672cagattacggggcetytactcattyagccattyacccactyagacaccatyagacac684ctcataggcttaagcaattttcaactcaatacttyacccactyagacaccatyagacac684ctcattyccttaagcaattttcaactcaatacttyagagaccactyagacaccatyagacaccatyagacacgacttytecttaagcaacttycccagggggagagycccacttytygdatyagacccatyaggacaccatyaggacaccactyagacaccactyaggacaccactyaggacaccactyagacaccagccytyte702gactycccaccccactygtttcttyagtygactactcacattcttytagagactactyacccactyatytecagcyccaty702ttyccccaccccactygettccttyagettytygattyccaccattytyty720ttygaggacagacattyatytagtyaggaty <td>acgactagct</td> <td>gcggtaacac</td> <td>cctcacatgt</td> <td>tacttgaagg</td> <td>cctctgcagc</td> <td>ctgtcgagct</td> <td>6360</td>	acgactagct	gcggtaacac	cctcacatgt	tacttgaagg	cctctgcagc	ctgtcgagct	6360
tactetgece cecegggga cecgeceaa ceagaataeg acttggaget gataacatea 654 tgttecteca attgtgeggt egecagaat geateaggea aaagggggta ctactectaec 660 egeaceggget gegtgggaaa caagetagaac acteceagt 666 aacteetgge taggeacaat tateatgtat gegteggaaa cagetagaaca acteceagt 666 aacteetgge taggeacaat tateatgtat gegtegaaaca ttggagaaag gatgattetg 672 atgacteact tettetecat ectetatgag caggagaac ttggagaaaag caggagaac caggagaaac ttgaaaaaga cetegagaatgact cecatggge tteaggaaaca tteeatgag cacttgace taceatgaga cattgaaca cattgaaca 684 acggaceagg caggagaaca ttggagaaaca ttggaggaaca cattgagaca caggagaaca ttggagcaaga gaggagaca cattgagac tecaatggg gaggaggecg cacttgtgg caatagagacaggactgactgaggagaacacagggggagagaacacaggaggagagacacaggagg	gcgaagctcc	aggactgcac	gatgctcgtg	aacgccgccg	gccttgtcgt	tatctgtgaa	6420
tactetgece cecegggga cecgeceaa ceagaataeg acttggaget gataacatea 654 tgttecteca attgtgeggt egecagaat geateaggea aaagggggta ctactectaec 660 egeaceggget gegtgggaaa caagetagaac acteceagt 666 aacteetgge taggeacaat tateatgtat gegteggaaa cagetagaaca acteceagt 666 aacteetgge taggeacaat tateatgtat gegtegaaaca ttggagaaag gatgattetg 672 atgacteact tettetecat ectetatgag caggagaac ttggagaaaag caggagaac caggagaaac ttgaaaaaga cetegagaatgact cecatggge tteaggaaaca tteeatgag cacttgace taceatgaga cattgaaca cattgaaca 684 acggaceagg caggagaaca ttggagaaaca ttggaggaaca cattgagaca caggagaaca ttggagcaaga gaggagaca cattgagac tecaatggg gaggaggecg cacttgtgg caatagagacaggactgactgaggagaacacagggggagagaacacaggaggagagacacaggagg	agcgcgggaa	cccaagagga	cgcggcgagc	ctacgagtct	tcacggaggc	tatgactagg	6480
cgttactcca atgtgteggt cgccacgat gcatcaggca aaagggtgta ctacctcacc 660 cgtgatcca cacccccct cgcacgggt gcgtgggaaa cagctagaca cactccagtt 666 aactcctggc taggcaacat tatcatgtat gcgccactt tgtgggcaag gatgattcg 672 atgactcacg gggcctgtta ctccttagac cacgtgacacat tgtgaacaagc cctggacttgc ctacgccattt ttcactccat agttactctc caggtgagaac cattgaacga 684 ctccatggcc ttaggcactt ttcactccat agttactctc caggtgagac caatagggtg gcttcatgc ctaggacact tggggtacc cccttgcag tctggagaca cattgaacga 690 gcttcatgcc ctaggaacat tggggtacc cccttgcagg tctggagaca cattgaacga 696 gcttcatgcc gctggttcgt tgctggggaca cccttgcag tctggagaca cccttgagg gcactgtgcg ctaggtact tgctgggtac aactcacaa cccaggtag tctgagagac caactcacat caggagagac ccacttgtgg cattgtcg gctggttcg tgctggttc catgcttgc ctactccaat cccggctgg tccaactgg ggagggggg acatatatca cagcctgtct tacctgctc ccaacggaa actcacacaa actcactcaat ctccaggtag ggtaggagact tggccagag ggaaggagggggggggg	tactctgccc	ccccgggga	cccgccccaa	ccagaatacg	acttggagct	gataacatca	6540
aactoctggc taggcacat tatcatgat gegeceactt tgtgggcaag gatgattetg 672 atgacteact tetteteat cettetagac caggagcaac ttgaaaaaagc cettgagactg cagatetaag gggeetgta etceattgag ceaettgace tacetcagat caattagacg ctcatggce ttagegcatt tteacteat agttactete caggtgagat caatagggtg getteatgce tcaggaaact tggggtacac ceettgegag tetggagaca teggggeeggg getggteegg ctaggetact geecaggg gggagggeeg caattagtg caatagggt getteatgge cagggaagac caaacteaaa ctcacaa teceggeegg tegggeeggg getggttegg getggttegt tgetggttac ageggggggag acatatatac caggeeggg getggeegge ceageggt aatetagage tgtgeegggg ggaagggeegg teggeegga ceegggggag acatatatac caggeeggg teggeegga ceeggggggag acatatatac caggeeggg teggeeggeegggag acateagag ggaagggeegggag acatatatac caggeeggg teggeeggeegggaga acateagagggggggggggggggggggggggggggggggg	tgttcctcca	atgtgtcggt	cgcccacgat	gcatcaggca	aaagggtgta	ctacctcacc	6600
atgactcact tettetecat cettetagea caggageaca ttgaaaaage cetggactge 678 cagatetacg gggcetgtta etccattgag ceacttgace tacctcagat ctccatggec ttaggceatt tteactcat agtactee caggtgagat getteatgeg ctaggacatt ttgaggacact cacattgagg ggagggggg cacattgtgg caatgagaca caaactcaaa ctcactcaa teceggetgc gtccaggg gactggagg acaataatac acagcetgtet tgtgecegac cecgetggtt catggttgac cacactcaa teceggetg ggaggggacact caggetget cacceggtgat catggttgag caactactaa accactcaact tttetgtagg ggagggggggggggggggggggggggg	cgtgatccca	ccacccccct	cgcacgggct	gcgtgggaaa	cagctagaca	cactccagtt	6660
atgactcacttettetcactcettetagecagagacaacttgaaaaagccettgaactgactcagtactgactgactgactgactcagtactgactgactgactgactgactgactgactgactg	aactcctggc	taggcaacat	tatcatgtat	gcgcccactt	tgtgggcaag	gatgattctg	6720
ctccatggcc ttagcgcatt ttcactcat agttactctc caggtgagat caatagggtg gettcatgcc tcaggaaact tggggtacca cccttgcgag tctggagaca tcgggcagg 696 agcgtcactgg caggagaca caaactcaaa ctcactcaa tcccggctgc gtccaggcg ftcactggg caggagaca caaactcaaa ctcactcaa tcccggctgc gtccaggctg tgtgtgtac agcgggggag acatatatac cagcctgtc 712 cgtgccgac cccgctggtt catgctgtgc ctactcctac ttctgtagg gtaggcatc 720 tacctgctc ccaaccggta aatctaaagg tgtgccttct agtgccctc ccacgggta cattaggac acttgaggggag acatatatac cagcctgtct 714 cgtgccctc ccaaccggta aatctaaagg tgtgccttct agttgccagc catcttgtgt ftgccgttc cgaagggggaggagggggggggggggggggg	atgactcact	tcttctccat	ccttctagca	caggagcaac	ttgaaaaagc	cctggactgc	6780
ctccatggcc tagggcatt ttcactcat agttactcc caggtgagat caatagggtg gettcatgcc tcaggaaact tggggtacca cccttgcgag tctggagaca tcgggcagg ggagggcccg caactctgtgg caagtacctc 702 ttcaactggg cagtgagaca caaactcaaa ctcactcaaa tccactcgaa tcccggctg getccagctg 708 gacttgtccg gctggttcgt tgctggttac agcggggggg acatatatca cagcctgtct 714 cggcccgac cccactggtt catgctgtgc ctcactccaa ttcctgtagg ggaggggag acatatatca cagcctgtct 714 cggccctcc ccaaccggta aactagagc tgtgccttct agttgccgcc ccaccggtg acatctgaact ggaaggtgc actccactgt tcctccacccactg tcctccact ttcctgtagg ggagggggggggg	cagatctacg	gggcctgtta	ctccattgag	ccacttgacc	tacctcagat	cattgaacga	6840
agcytcogog ctaggotact gtccagagg ggagaggocg ccacttytgg caagtacetc 702 ttcaactggg caytgaagac caaactcaaa ctcactcaa tcccgctgc gtccaagctg 708 gacttytcog gctgyttogt tgctggtac agcgggggag acatatatca cagcctytct 714 cgtgoccgac cccgctggt catgotggc ctactcctac ttcctgagg ggagaggatc 720 tacctgctcc ccaaccggta aatctagagc tgtgccttc agttgccgcg cccgctggt catgotggc ctactcctac tttctgtagg ggagagcatc 720 tacctgctcc ccaaccggta aatctagagc tgtgccttc agttgccagc ctactcgttct agttgccagc actctgttgt tggagagagatgg ggagaggatgg ggagaggatgg ggagagaga	ctccatggcc	ttagcgcatt	ttcactccat	agttactctc	caggtgagat	caatagggtg	6900
ttcaactggg cagtgaagac caaactcaaa ctcactcaa tcccggctgc gtcccagctg gacttgtccg gctggttcgt tgctggttac agcgggggag acatatatca cagcctgtct tacctgctcc ccaaccggta aatctagagc tgtgccttct agttgccagc catctgttgt ttgcccccc ccaaccggta aatctagagc tgtgccttct agttgccagc catctgttgt ttgcccccc cccgtgctt ccttgaccct ggaaggtgcc actcccactg tcctttctaa ttaaaatgag gaaattgcat cgcattgtct gagtaggtgc actccactg tcctttctaa taaaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattc tggggggtgg gaggagggggggggg	gcttcatgcc	tcaggaaact	tggggtacca	cccttgcgag	tctggagaca	tcgggccagg	6960
gacttgtccg gctggttcgt tgctggttaa agcgggggag acatatataca cagcctgtct catgccgcc cccaccggta aatctagagc tgtgccttct agttgccagc catctgttgt 726 ttgccctcc cccagtgctt ccttgaccct ggaaggtgc actcccactg tccttctcta ataaaatgag gaaattgcat cgcattgtct gagtaggtgc actcccactg tccttctcta ataaaatgag gacagcaggaggaggaggaggaggaggaggaggaggagga							7020
cgtgcccgac cccgctggtt catgctgc ctactcctac tttctgtagg ggtaggcatc 720 tacctgctcc ccaaccggta aatctagagc tgtgccttct agttgccagc catctgttgt 726 ttgccctcc cccgtgcctt ccttgaccct ggaaggtgc actccactg tcctttccta ataaatgag gaaattgcat cgcattgtct gagtaggtg cattctattc tggggggtgg 738 ggtggggcag gacagcaagg gggaggattg ggaagacaat actgagggtgg ggaagacat atggccgac ggcgccgt actgagggtg cattctattc tggggggtgg 738 ggtgggggcd actgagggtg cattcattc tggggggtgg 738 ggtggggcct atggccgac ggcgccgt actgagagtg cttaaggggtgg cttaaggggtgg cttaagggtgggggggggg							7080
tacctgctcc ccaaccggta aatctagagc tgtgcttct agttgccagc catctgttgt tggccctcc cccgtgctt ccttgacct ggaaggtgc actccactg tcctttcta agttgggggggggg	gacttgtccg	gctggttcgt	tgctggttac	agcgggggag	acatatatca	cagcctgtct	7140
tacctgctcc ccaaccggta aatctagagc tgtgccttct agttgccagc catctgttgt ttgccctcc cccgtgcctt ccttgaccct ggaaggtgcc actcccactg tccttccta 732 ataaaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattc tggggggtgg ggagggggggggg	cgtgcccgac	cccgctggtt	catgctgtgc	ctactcctac	tttctgtagg	ggtaggcatc	7200
ataaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattc tgggggtgg 738 ggtggggcag gacagcaagg gggaggattg ggaagacaat agcaggcatg ctggggatgc 744 ggtgggctct atggccgatc ggcgccgt actgaaatgt gtgggcgtgg cttaagggtg 750 ggaaagaata tataaggtgg gggtcttatg tagttttgta tctgttttgc agcagccgcc 756 gccgccatga gcaccaactc gtttgatga agcattgtga gctcatattt gacaacgcgc 756 gtcctgcccq caaactctac taccttgacc tacgagaccg tgtctggaac gccgttggag 774 actgcagcct ccgccgcgc ttcagccgc gcagccaccg cccgcgggat ttcagccgct tcaggaccg tccggggat gcgtcagaat gtgatgggct ccagcattga tggtcgccc 768 gtctgctcc tgaacccgc ttcagccgct tgcaagcacg gcagccaccg cccgcgggat tgtgactgac 774 actgcatcc tacgagccgt tcagccgct tcaggaccg tccgggggat gcgagccaccg gcagccaccg cccgcgggat tgtgactgac 774 actgctct tgaacccgct tgcaagcagt gcagccaccg gcagcaccg cccgcggggat tgtgactgac 774 actgctctc tgaacccgc tcagccgct tcaggaccg gcagccaccg cccgcggggat tgtgactgac 774 actgctctc tgaacccgct tgcaagcagt gcagccaccg gcagcaccg gtcatccgc ccgcgagag 774 actgctctc tgaacccgc tcagaccgc gcagccaccg gcagcaccg gccgttggag 774 actgctctc tgaacccgc tcagaccgc gcagccaccg gcagcaccg gccggatgac 774 actgctcc tgaacccgc tcagaccacc gcagcaggat tctgcgccc gggaaccac gccgcgggatgac 774 actgctgccc caacctcaccgcgcgggtttggacgcccc tggaaccacc gcagcagggt tctgcgccc gggaaccacc gccgcggggttgaccacactac gggagaccaccac gggaaccaccac gccggatgaccaccac gcagaggttc actgcggtct actgcgggggggggg	tacctgctcc	ccaaccggta	aatctagagc	tgtgccttct	agttgccagc	catctgttgt	7260
ataaaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattc tgggggtgg gagtgggggggggg	ttgcccctcc	cccgtgcctt	ccttgaccct	ggaaggtgcc	actcccactg	tcctttccta	7320
ggtgggctct atggccgatc ggcgcgcgt actgaaatgt gtgggcgtgg cttaaggtg 750 ggaaaagaata tataaggtgg gggtcttatg tagttttgta tctgttttgc agcaaccgc 756 gccgccatga gcaccaactc gtttgatgga agcattgtga gctcatattt gacaaccgcg 762 atgcccccat gggccggggt gcgtcagaat gtgatgggct ccagcattga tggtcgccc 768 gtcctgcccg caaactctac taccttgacc tacgagaccg tgtctggaac gccgttggag 774 actgcagcct ccgccgcgc ttcagccgct gcagcaccg cccgcgggat tgtgactgac 780 tttgctttcc tgagccgct tgcaagcagt gcagcttccc gttcatccgc ccgcgatgac 786 aagttgacgg ctcttttggc acaattggat tcttgaccc gggaacttaa tgtcgttct 792 cagcagctgt tggatctgcg ccagcaggtt tctgccctga aggcttcctc ccctccaat 798 gcggtttaaa acataaataa aaaaccagac tctgtttgga tttggatcaa gcagtgtct tgcagggcc cggtaggcc gggaccaccg gcgaactaa gccgtctt tttccaggacg tggtaaaggt gactctctga gttcagcag gttcggtcg ttggaggtc tggtaaaggt gactctggt gttcagatac 816 atgggcataa gccgtctct gggggtggagg tagcaccact gcagagctt aagacgggtgggggggggg	ataaaatgag	gaaattgcat	cgcattgtct	gagtaggtgt	cattctattc	tggggggtgg	7380
ggaaagaata tataaggtgg gggtcttatg tagttttgta tctgttttgc agcagccgcc 756 gccgccatga gcaccaactc gtttgatgga agcattgtga gctcatattt gacaacgcgc 762 atgccccat gggccggggt gcgtcagaat gtgatgggct ccagcattga tggtcgcccc 768 gtcctgcccg caaactctac taccttgacc tacgagaccg tgtctggaac gccgttggag 774 actgcagcct ccgccgccgc ttcagccgct gcagccaccg cccgcgggat tgtgactgac 780 tttgctttcc tgagcccgct tgcaagcagt gcagctccc gttcatccgc ccgcgatgac 786 aagttgacgg ctcttttggc acaattggat tctttgaccc gggaacttaa tgtcgttct 792 cagcagctgt tggatctgcg ccagcaggt tctgcctga aggcttcctc ccctcccaat 798 gcggtttaaa acataaataa aaaaccagac tctgtttgga tttggatcaa gcaagtgtc tttggagggtcc tgtgtattt ttccaggacg tggtagagg gggaccagcg gtctcggtcg 810 ttgagggatca agccgtctct ggggtggagg tagcaccact gcagagctt atgctgcggg 822 gtgtgtgttg agatagacca gtcgtagag gagcaccact gcagagctta aaaatgtct agctggatg ggtgcatacg ggggagag cccttggtg agattgtcc cagcagaac tcggggatg aggcaccac gcggggttga agctgggatg ggtgcatacg tggggatatg agatgaccac tgggatgtta aaaatgtct agctgggatg ggtgcatacg tggggatatg agatgcatct tggacctga aaaatgtct agctgggatg ggtgcatacg tggggatatg agatgcatct tggacctga aaaaatgtct agctgggat ggtgcatacc cccccgggga ttcatgttg gcagaaccac cagcacagtg 846 tatccggtgc acttgggaa tttgtcatgt agcttagaa ggaaatttc agcatagtgc gaagaacttg ggaagacccc tgtggac acttggga gaagatatt ctgggatcac taacgtcata gttgttcc aggaccgcg cggcctggg cggcctggg cggcctggg gaagaacttg ggaagaccac catttttaca aagcgggg ggaagaccac agcacagtg 858 ccacgggcgg cggcctggg cggcctggg catttttaca aagcgcggg ggaagatcac agcacggt 870 agaagatgagat cgtcataggc catttttaca aagcgcggg ggaagatgcc agaactgcgt 870 agaactgcg ggaagatgcc agaagatgcc agaagatgcc agaagatttcc agaagatgcc agaagatttcc agaagatgcc agaagatgcc agaagatgcgc agaagactgcg agaagatgcgc agaagatgcgc agaagatgcgc	ggtggggcag	gacagcaagg	gggaggattg	ggaagacaat	agcaggcatg	ctggggatgc	.7440
ggaaagaata tataaggtgg gggtcttatg tagttttgta tctgttttgc agcagccgcc 756 gccgccatga gcaccaactc gtttgatgga agcattgtga gctcatattt gacaacgcgc 762 atgccccat gggccggggt gcgtcagaat gtgatgggct ccagcattga tggtcgcccc 768 gtcctgcccg caaactctac taccttgacc tacgagaccg tgtctggaac gccgttggag 774 actgcagcct ccgccgccgc ttcagccgct gcagccaccg cccgcgggat tgtgactgac 780 tttgctttcc tgagcccgct tgcaagcagt gcagctccc gttcatccgc ccgcgatgac 786 aagttgacgg ctcttttggc acaattggat tctttgaccc gggaacttaa tgtcgttct 792 cagcagctgt tggatctgcg ccagcaggt tctgcctga aggcttcctc ccctcccaat 798 gcggtttaaa acataaataa aaaaccagac tctgtttgga tttggatcaa gcaagtgtc tttggagggtcc tgtgtattt ttccaggacg tggtagagg gggaccagcg gtctcggtcg 810 ttgagggatca agccgtctct ggggtggagg tagcaccact gcagagctt atgctgcggg 822 gtgtgtgttg agatagacca gtcgtagag gagcaccact gcagagctta aaaatgtct agctggatg ggtgcatacg ggggagag cccttggtg agattgtcc cagcagaac tcggggatg aggcaccac gcggggttga agctgggatg ggtgcatacg tggggatatg agatgaccac tgggatgtta aaaatgtct agctgggatg ggtgcatacg tggggatatg agatgcatct tggacctga aaaatgtct agctgggatg ggtgcatacg tggggatatg agatgcatct tggacctga aaaaatgtct agctgggat ggtgcatacc cccccgggga ttcatgttg gcagaaccac cagcacagtg 846 tatccggtgc acttgggaa tttgtcatgt agcttagaa ggaaatttc agcatagtgc gaagaacttg ggaagacccc tgtggac acttggga gaagatatt ctgggatcac taacgtcata gttgttcc aggaccgcg cggcctggg cggcctggg cggcctggg gaagaacttg ggaagaccac catttttaca aagcgggg ggaagaccac agcacagtg 858 ccacgggcgg cggcctggg cggcctggg catttttaca aagcgcggg ggaagatcac agcacggt 870 agaagatgagat cgtcataggc catttttaca aagcgcggg ggaagatgcc agaactgcgt 870 agaactgcg ggaagatgcc agaagatgcc agaagatgcc agaagatttcc agaagatgcc agaagatttcc agaagatgcc agaagatgcc agaagatgcgc agaagactgcg agaagatgcgc agaagatgcgc agaagatgcgc	ggtgggctct	atggccgatc	ggcgcgccgt	actgaaatgt	gtgggcgtgg	cttaagggtg	7500
atgccccat gggcggggt gcgtcagaat gtgatgggct ccagcattga tggtcgccc 768 gtcctgcccg caaactctac taccttgacc tacgagaccg tgtctggaac gccgttggag 774 actgcagcct ccgcgcgcc ttcagccgct gcagcaccg ccgcgggat tgtgactgac 780 tttgctttcc tgagccgct tgcaagcagt gcagcttccc gttcatccgc ccgcgatgac 786 aagttgacgg ctcttttggc acaattggat tctttgaccc gggaacttaa tgtcgttct 792 cagcagctgt tggatctgcg ccagcaggt tctgccctga aggcttcctc ccctccaat 798 gcggtttaaa acataaataa aaaaccagac tctgtttgga tttggatcaa gcaagtgtct gcggtttct atttaggggt tttgcgccgc cggtaggccc gggaaccagcg gtctcggtcg 810 ttgagggtcc tgtgatttt ttccaggacg tggtaaaggt gactctggat gttcagatac 816 atgggcataa gcccgtctt ggggtggagg tagcaccact gcagagcttc atgctggg 822 gtggtgttgt agatgatca gtcgtagcag gacgctggg cgtggtgcct aaaaatgct 828 ttcagtagca agctgattgc caggggatag agatgcatct tggactgat ttttaggttg gactagttac 834 agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgat ttttaggttg gaagaccac cagcacagtg gaagaccac tgtggtgc acttgggaa tttgtcatgt gcagaaccac cagcacagtg gaagacgcct tgtgacctc aagatttcc atgcattcgt ccataatgat ggcaatggc gaagacccc tgtgacctc aagattttcc atgcattcgt ccataatgat ggcaatggc gaagaccac aggacgccc aggacgccc caccggcgg cggcctggg cggctggg cggcataggc ggaagacacc aggacaggc 858 ccacgggcgg cggcctgggc catttttaca aagcgcggc ggaaggtgcc agaactcc agacctggc agaagatatt ctgggatcac taacgtcata gttgtgtcc aggaagacac catttttaca aagcgcggc ggaagggtgcc agaactcc agacctggc agaagatatt ctgggatcac taacgtcata gttgtgtcc aggaagacgcc aggacggcg catttttaca aagcgcggc ggaagggcgc agaactcc agacctggc gaagatattt ctgggatcac taacgtcata gttgtgtcc aggaagacgcc aggactgggc catttttaca aagcgcggg ggaagggtgcc agaactcc agacctgggc agaagatgcc agaagacacac agacctgggc agaagatgcc agaagacacac agaccacacacacacacacacaca	ggaaagaata	tataaggtgg	gggtcttatg	tagttttgta	tctgttttgc	agcagccgcc	7560
gtcctgccg caaactctac taccttgacc tacgagaccg tgtctggaac gccgttggag 774 actgcagcct ccgccgcgc ttcagccgct gcagccaccg cccgcgggat tgtgactgac 780 tttgcttcc tgagcccgct tgcaagcagt gcagcttccc gttcatccgc ccgcgatgac 786 aagttgacgg ctcttttggc acaattggat tctttgaccc gggaacttaa tgtcgttct 792 cagcagctgt tggatctgcg ccagcaggtt tctgccctga aggcttcctc cccccaat 798 gcggtttaaa acataaataa aaaaccagac tctgtttgga tttggatcaa gcaagtgtct gtgagggtcc tgtgtattt ttccaggacg cggtaggccc gggaccagcg gtctcggtcg 810 ttgagggtcc tgtgtatttt ttccaggacg tggtaaaggt gactctggat gttcagatac 822 gtggtgttgt agatgatcca gtcgtagcag gagcgctgg cgtggtgcct aaaaatgtct 828 ttcagtagca agctgattgc caggggcagg cccttggtgt aagtgtttac aaaactgtct 828 ttcagtagca agctgattgc caggggcagg cccttggtgt aagtgtttac aaaacgggtta 840 gctatgttcc cagccatacc cctccggga ttcatgttgt gcagaaccac cagcacagtg 846 taccggtgc acttgggaaa tttgtcatgt agcttagaag gaaatgcgtg gaagaacttg 826 gagacgccct tgtgacctcc aagattttcc atgcatcgt ccataatgat ggcaatgggc 8858 ccacgggcgg cggcctgggc gaagatattt ctgggatcac taacgtcata gttgtgttcc 8644 aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcgt 870	gccgccatga	gcaccaactc	gtttgatgga	agcattgtga	gctcatattt	gacaacgcgc	7620
actgcagcct cegecgeege tteagecget geagecaceg ceegegggat tgtgactgae 780 tttgetttee tgageceget tgeaageagt geagetteee gtteateege cegegatgae 786 aagttgaegg etetttgge acaattggat tettgaeee gggaaettaa tgtegtteet 792 cageagetg tggatetgee eageaget tettgeeetga aggetteete eecteeaat 798 geggtttaaa acataaataa aaaceagae tetgtttgga tttggateaa geaagtgeet tgegggeege eggtaggee eggetggeegggeeg	atgcccccat	gggccggggt	gcgtcagaat	gtgatgggct	ccagcattga	tggtcgcccc	7680
tttgctttcc tgagcccgct tgcaagcagt gcagcttccc gttcatccgc ccgcgatgac 786 aagttgacgg ctcttttggc acaattggat tctttgaccc gggaacttaa tgtcgtttct 792 cagcagctgt tggatctgcg ccagcaggtt tctgccctga aggcttcctc ccctccaat gcggtttaaa acataaataa aaaccagac tctgtttgga tttggatcaa gcaagtgtct tttgagggtcc tgtgtattt ttccaggacg cggtaggcc gggaccagcg gtctcggtcg tttgagggtcc tgtgtattt ttccaggacg tggtaaaggt gactctggat gttcagatac 816 atgggcataa gcccgtctct ggggtggagg tagcaccact gcagagcttc atgctgcggg gtggggtgttgt agatgatcca gtcgtagcag gacgctggg cgtggtgcct aaaaatgtct 828 ttcagtagca agctgattgc caggggcagg cccttggtgt aagtgtttac aaagcggtta 826 agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgtat ttttaggttg gcagagctc cagccatatc cctccggga ttcatgttgt gcagaaccac cagcacagtg 846 accggtgc acttgggaa tttgtcatgt agcttagaag gaaatgcgtg ggaagaacttg ggcaatggc cagcctggg cggctggg cgaagagccc tgtgacccc aagattttcc atgcatcgt ccataatgat ggcaatgggc gaagatatt ctgggatca taacgtcata gttgtgtcc aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcgt 8864 aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcgt 8870 aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcgt 8870 aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcgt	gtcctgcccg	caaactctac	taccttgacc	tacgagaccg	tgtctggaac	gccgttggag	7740
aagttgacgg ctcttttggc acaattggat tctttgaccc gggaacttaa tgtcgtttct 792cagcagctgt tggatctgcg ccagcaggtt tctgccctga aggcttcctc ccctccaat gcggtttaaa acataaataa aaaaccagac tctgtttgga tttggatcaa gcaagtgtct tttgagggcc gggtaggcc gggaccagcg gtctcggtcg 810ctgagggcc tgtgtattt ttccaggacg tggtaaaggt gactctggat gttcagatac 822cgggtgttgt agatgatcca gtcgtagcag gagggctggg cgtggtgcc aaaaatgtct 82ggggtgttgt agatgatcca gtcgtagcag gagggctggg cgtggtgcct aaaaatgtct 82gggtggatg ggtgagatg ggtgcatacg tggggatatg agatgatca 82gggatggatg ggtgagatg ggtgagatgatg cccttggtgt aagtgtttac aaaagcggtta 82ggctgggatggatggatggatggatggatggatggatgga	actgcagcct	ccgccgccgc	ttcagccgct	gcagccaccg	cccgcgggat	tgtgactgac	7800
cagcagctgt tggatctgg ccagcaggtt tctgcctga aggcttcctc ccctccaat 798 gcggtttaaa acataaataa aaaaccagac tctgtttgga tttggatcaa gcaagtgtct tgctgtctt atttaggggt tttgcggcg cggtaggcc gggaccagcg gtctcggtcg 810 ttgagggtcc tgtgtattt ttccaggacg tggtaaaggt gactctggat gttcagatac 816 atgggcataa gcccgtctct ggggtggagg tagcaccact gcagagcttc atgctgcgg 822 gtggtgttgt agatgatcca gtcgtagcag gagcgctggg cgtggtgcct aaaaatgtct 828 ttcagtagca agctgattgc caggggcagg cccttggtgt aagtgtttac aaagcggtta 826 agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgtat ttttaggttg 840 gctatgttcc cagccatatc cctccggga ttcatgttgt gcagaaccac cagcacagtg 846 tatccggtgc acttgggaaa tttgtcatgt agcttagaag gaaatgcgtg gaagaacttg 852 gagacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc 858 ccacgggcgg cggcctgggc gaagatattt ctgggatcac taacgtcata gttgtgtcc 864 aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcgt 870 aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcggt	tttgctttcc	tgagcccgct	tgcaagcagt	gcagcttccc	gttcatccgc	ccgcgatgac	
gcggtttaaa acataaataa aaaaccagac tctgtttgga tttggatcaa gcaagtgtct tgctgtctt atttaggggt tttgcgcgc cggtaggccc gggaccagcg gtctcggtcg tttgagggtcc tgtgtattt ttccaggacg tggtaaaggt gactctggat gttcagatac atgggcataa gcccgtctct ggggtggagg tagcaccact gcagagcttc atgctgcggg gtgtgttgt agatgatcca gtcgtagcag gagcgctggg cgtggtgcct aaaaatgtct ttcagtagca agctgattgc caggggcagg cccttggtgt aagtgttac aaaagcggtta agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgtat ttttaggttg gctatgttcc cagccatatc cctccggga ttcatgttgt gcagaaccac cagcacagtg gagacgccct tgtgacctcc aagattttcc atgcattcgt gcagaaccac cagcacagtg gagacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc gagacggcg cggctgggc gaagatatt ctgggatcac taacgtcata gttgtgtcc aggatgagat cgtcataggc cattttaca aagcgcggc ggagggtgcc agactgcgt 8706	aagttgacgg	ctcttttggc	acaattggat	tctttgaccc	gggaacttaa	tgtcgtttct	
tgctgtcttt atttaggggt tttgcgccg cggtaggccc gggaccagcg gtctcggtcg ttgagggtcc tgtgtatttt ttccaggacg tggtaaaggt gactctggat gttcagatac atgggcataa gcccgtctct ggggtggagg tagcaccact gcagagcttc atgctgcggg gtggtgttgt agatgatcca gtcgtagcag gagcgctggg cgtggtgcct aaaaatgtct agctgggatg ggtgcatacg caggggcagg cccttggtgt aagtgtttac aaagcggtta agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgtat ttttaggttg gctatgttcc cagccatatc cctccgggga ttcatgttgt gcagaaccac cagcacagtg gagacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc gaagatattt ctgggatcac taacgtcata gttgtgtcc aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcgt 8840	cagcagctgt	tggatctgcg	ccagcaggtt	tctgccctga	aggcttcctc	ccctcccaat	
ttgagggtcc tgtgtatttt ttccaggacg tggtaaaggt gactctggat gttcagatac atgggcataa gcccgtctct ggggtggagg tagcaccact gcagagcttc atgctgcggg gtggtgttgt agatgatcca gtcgtagcag gagcgctggg cgtggtgcct aaaaatgtct ttcagtagca agctgattgc caggggcagg cccttggtgt aagtgttac aaagcggtta agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgtat ttttaggttg gctatgttcc cagccatatc cctccggga ttcatgttgt gcagaaccac cagcacagtg tatccggtgc acttgggaaa tttgtcatgt agcttagaag gaaatgcgtg gaagaacttg ggaacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc gaagatggc catttttaca aagcgcggc ggagggtgcc agactgcgt sactggga gaagatattt ctgggatcac taacgtcata gttgtgtcc agactgcgc catttttaca aagcgcggc ggagggtgcc agactgcgt 870	gcggtttaaa	acataaataa	aaaaccagac	tctgtttgga	tttggatcaa	gcaagtgtct	8040
atgggcataa gcccgtctct ggggtggagg tagcaccact gcagagcttc atgctgcggg gtggtgttgt agatgatcca gtcgtagcag gagcgctggg cgtggtgcct aaaaatgtct tcagtagca agctgattgc caggggcagg cccttggtgt aagtgttac aaaagcggtta agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgtat ttttaggttg gctatgttcc cagccatatc cctccggga ttcatgttgt gcagaaccac cagcacagtg gagacgccct tgtgacctcc aagattttcc atgcattcgt gcagaaccac cagcacagtg gagacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc gagacggcg cggcctgggc gaagatattt ctgggatcac taacgtcata gttgtgtcc aggatgatgat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcggt 870	tgctgtcttt	atttaggggt	tttgcgcgcg	cggtaggccc	gggaccagcg	gtctcggtcg	8100
gtggtgttgt agatgatcca gtcgtagcag gagcgctggg cgtggtgcct aaaaatgtct tcagtagca agctgattgc caggggcagg cccttggtgt aagtgtttac aaagcggtta agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgtat ttttaggttg gctatgttcc cagccatatc cctccgggga ttcatgttgt gcagaaccac cagcacagtg gagacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc gaagatattt ctgggatcac taacgtcata gttgtgtcc aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcgt 870	ttgagggtcc	tgtgtatttt	ttccaggacg	tggtaaaggt	gactctggat	gttcagatac	
ttcagtagca agctgattgc caggggcagg cccttggtgt aagtgtttac aaagcggtta 834 agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgtat ttttaggttg gctatgttcc cagccatatc cctccgggga ttcatgttgt gcagaaccac cagcacagtg gagacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc gaagatattt ctgggatcac taacgtcata gttgtgttcc aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcgt 870							
agctgggatg ggtgcatacg tggggatatg agatgcatct tggactgtat ttttaggttg gctatgttcc cagccatatc cctccgggga ttcatgttgt gcagaaccac cagcacagtg tatccggtgc acttgggaaa tttgtcatgt agcttagaag gaaatgcgtg gaagaacttg ggaacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc cacagggggg cggcctgggc gaagatattt ctgggatcac taacgtcata gttgtgttcc aggatgatgat cgtcataggc catttttaca aagcgcgggc ggagggtgcc agactgcggt 8700							
gctatgttcc cagccatatc cctccgggga ttcatgttgt gcagaaccac cagcacagtg tatccggtgc acttgggaaa tttgtcatgt agcttagaag gaaatgcgtg gaagaacttg gagacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc ccacgggcgg cggcctgggc gaagatattt ctgggatcac taacgtcata gttgtgttcc aggatgagat cgtcataggc catttttaca aagcgcgggc ggagggtgcc agactgcggt 846 852 864 870	ttcagtagca	agctgattgc	caggggcagg	cccttggtgt	aagtgtttac	aaagcggtta	
tatccggtgc acttgggaaa tttgtcatgt agcttagaag gaaatgcgtg gaagaacttg gagacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc ccacgggcgg cggcctgggc gaagatattt ctgggatcac taacgtcata gttgtgttcc aggatgagat cgtcataggc catttttaca aagcgcgggc ggagggtgcc agactgcggt 870	agctgggatg	ggtgcatacg	tggggatatg	agatgcatct	tggactgtat	ttttaggttg	8400
gagacgccct tgtgacctcc aagattttcc atgcattcgt ccataatgat ggcaatgggc 858 ccacgggcgg cggcctgggc gaagatattt ctgggatcac taacgtcata gttgtgttcc aggatgagat cgtcataggc catttttaca aagcgcggc ggagggtgcc agactgcggt 870	gctatgttcc	cagccatatc	cctccgggga	ttcatgttgt	gcagaaccac	cagcacagtg	8460
ccacgggcgg cggcctgggc gaagatattt ctgggatcac taacgtcata gttgtgttcc 8640 aggatgagat cgtcataggc catttttaca aagcgcgggc ggagggtgcc agactgcggt 8700	tatccggtgc	acttgggaaa	tttgtcatgt	agcttagaag	gaaatgcgtg	gaagaacttg	
aggatgagat cgtcataggc cattttaca aagcgcgggc ggagggtgcc agactgcggt 870							8580
- maganagas							8640
ataatggttc catccggccc aggggcgtag ttaccctcac agatttgcat ttcccacgct 876	aggatgagat	cgtcataggc	catttttaca	aagcgcgggc	ggagggtgcc	agactgcggt	8700
	ataatggttc	catccggccc	aggggcgtag	ttaccctcac	agatttgcat	ttcccacgct	8760

ttasattasa	atggggggat	catototaco	tacanaacaa	tgaagaaaac	gatttccaaa	8820
ctgagttcag	tcagctggga	agaaagaag	ttcctgagga	actacaactt	accacaacca	8880
gtagggaga	ccagciggga	tattaccocc	tacaactaat	anttagnana	actacaacta	8940
gragacecar	aaatcacacc	caccaccygc	ttaagatat	coctaactca	catattttcc	9000
ccgtcatccc	tgagcagggg	ggccactteg	ccaagcatge	ggagttettg	caegceecce	9060
ctgaccaaat	ccgccagaag	gegetegeeg	cccaycyata	ttttaaaaat	ttaaccaacc	9120
aagtttttca	acggtttgag	accgtccgcc	gtaggcatgc	t-t-agegt	ttgaccaage	9180
agttccaggc	ggtcccacag	ctcggtcacc	tgetetaegg	catctcgatc	cagcatatet	9240
cctcgtttcg	cgggttgggg	cggctttcgc	tgtacggcag	tagteggtge	tegtecagae	9300
gggccagggt	catgtctttc	cacgggcgca	gggtcctcgt	cagcgtagtc	tgggtcacgg	
tgaaggggtg	cgctccgggc	tgcgcgctgg	ccagggtgcg	cttgaggctg	greergergg	9360
tgctgaagcg	ctgccggtct	tegeeetgeg	cgtcggccag	gtagcatttg	accatggtgt	9420
catagtccag	ccctccgcg	gcgtggccct	tggcgcgcag	cttgcccttg	gaggaggcgc	9480
cgcacgaggg	gcagtgcaga	cttttgaggg	cgtagagctt	gggcgcgaga	aataccgatt	9540
ccggggagta	ggcatccgcg	ccgcaggccc	cgcagacggt	ctcgcattcc	acgagecagg	9600
tgagctctgg	ccgttcgggg	tcaaaaacca	ggtttccccc	atgctttttg	atgcgtttct	9660
tacctctggt	ttccatgagc	cggtgtccac	gctcggtgac	gaaaaggctg	tccgtgtccc	9720
cgtatacaga	cttgagaggc	ctgtcctcga	geggtgttcc	gcggtcctcc	tcgtatagaa	9780
actcggacca	ctctgagacg	aaggctcgcg	tccaggccag	cacgaaggag	gctaagtggg	9840
aggggtagcg	gtcgttgtcc	actagggggt	ccactcgctc	cagggtgtga	agacacatgt	9900
cgccctcttc	ggcatcaagg	aaggtgattg	gtttataggt	gtaggccacg	tgaccgggtg	9960
ttcctgaagg	ggggctataa	aagggggtgg	gggcgcgttc	gtcctcactc	tcttccgcat	10020
cgctgtctgc	gagggccagc	tgttggggtg	agtactccct	ctcaaaagcg	ggcatgactt	10080
ctgcgctaag	attgtcagtt	tccaaaaacg	aggaggattt	gatattcacc	tggcccgcgg	10140
tgatgccttt	gagggtggcc	gcgtccatct	ggtcagaaaa	gacaatcttt	ttgttgtcaa	10200
acttaataac	aaacgacccg	tagagggcgt	tggacagcaa	cttggcgatg	gagcgcaggg	10260
tttaatttt	gtcgcgatcg	gcgcgctcct	tggccgcgat	gtttagctgc	acgtattcgc	10320
gcgcaacgca	ccgccattcg	ggaaagacgg	tggtgcgctc	gtcgggcact	aggtgcacgc	10380
gccaaccgcg	gttgtgcagg	gtgacaaggt	caacgctggt	ggctacctct	ccgcgtaggc	10440
actcattaat	ccagcagagg	cggccgccct	tgcgcgagca	gaatggcggt	agtgggtcta	10500
actacatete	gtccgggggg	tctqcqtcca	cggtaaagac	cccgggcagc	aggcgcgcgt	10560
cgaagtagtc	tatcttgcat	ccttgcaagt	ctagcgcctg	ctgccatgcg	cgggcggcaa	10620
acacacactc	gtatgggttg	agtgggggac	cccatggcat	ggggtgggtg	agcgcggagg	10680
catacatacc	gcaaatgtcg	taaacgtaga	ggggctctct	gagtattcca	agatatgtag	10740
ggtagcatct	tccaccgcgg	atgctggcgc	gcacgtaatc	gtatagttcg	tgcgagggag	10800
cgaggaggtc	gggaccgagg	ttgctacggg	cgggctgctc	tgctcggaag	actatctgcc	10860
tgaagatggc	atgtgagttg	gatgatatgg	ttggacgctg	gaagacgttg	aagctggcgt	10920
ctgtgagacc	taccgcgtca	cgcacgaagg	aggcgtagga	gtcgcgcagc	ttgttgacca	10980
actcaacaat	gacctgcacg	tctagggcgc	agtagtccag	ggtttccttg	atgatgtcat	11040
acttatcctq	tccctttttt	ttccacagct	cgcggttgag	gacaaactct	tcgcggtctt	11100
tccagtactc	ttggatcgga	aacccgtcgg	cctccgaacg	gtaagagcct	agcatgtaga	11160
actogttgac	ggcctggtag	gcgcagcatc	ccttttctac	gggtagcgcg	tatgcctgcg	11220
caacetteea	gagcgaggtg	tagataaaca	caaaggtgtc	cctaaccatg	actttgaggt	11280
actogtattt	gaagtcagtg	tcatcacatc	cgccctgctc	ccagagcaaa	aagtccgtgc	11340
actititaga	acgcgggttt	ggcaggggga	aggtgacatc	gttgaagagt	atctttcccg	11400
cacaaaacat	aaagttgcgt	gtgatgcgga	agggtcccgg	cacctcggaa	cggttgttaa	11460
ttacctgggc	ggcgagcacg	atctcqtcaa	agccgttgat	gttgtggccc	acaatgtaaa	11520
gttccaagaa	gcgcgggatg	cccttgatgg	aaggcaattt	tttaagttcc	tcgtaggtga	11580
actetteaga	ggagctgagc	ccatactcta	aaagggccca	gtctgcaaga	tgagggttgg	11640
aadcdacdaa	tgagctccac	aggtcacggg	ccattagcat	ttgcaggtgg	tcgcgaaagg	11700
tectasacta	gcgacctatg	gccatttttt	ctaggataat	gcagtagaag	gtaagcgggt	11760
cttattccca	gcggtcccat	ccaaggtccg	caactaaatc	tegegegaca	gtcactagag	11820
actestetes	gccgaacttc	atdaccadca	tgaagggcac	gagetoette	ccaaaggccc	11880
ccatccaact	ataggtctct	acatentann	taacaaaaa	acactcaata	cgaggatgcg	11940
ageografice	gaagaactgg	atctccccc	accanttona	agataacta	ttgatgtggt	12000
agecyategy	gtccctgcga	Cadaccassc	actortocto	acttttataa	aaacatacac	12060
yaaaytayaa	guculyuga	cygyccyaac	accegugetg	500000gcaa		

agtactggca	gcggtgcacg	ggctgtacat	cctgcacgag	gttgacctga	cgaccgcgca	12120
caaggaagca	gagtgggaat	ttgagcccct	cgcctggcgg	gtttggctgg	tggtcttcta	12180
cttcggctgc	ttgtccttga	ccgtctggct	gctcgagggg	agttacggtg	gatcggacca	12240
ccacgccgcg	cgagcccaaa	gtccagatgt	ccgcgcgcgg	cggtcggagc	ttgatgacaa	12300
catcgcgcag	atgggagctg	tccatggtct	ggagctcccg	cggcgtcagg	tcaggcggga	12360
getectgeag	gtttacctcg	catagccggg	tcagggcgcg	ggctaggtcc	aggtgatacc	12420
tgatttccag	gggctggttg	gtggcggcgt	cgatggcttg	caagaggccg	catccccgcg	12480
gcgcgactac	ggtaccgcgc	ggcgggcggt	gggccgcggg	ggtgtccttg	gatgatgcat	12540
ctaaaagcgg	tgacgcgggc	gggcccccgg	aggtaggggg	ggctcgggac	ccgccgggag	12600
agggggcagg	ggcacgtcgg	cgccgcgcgc	gggcaggagc	tggtgctgcg	cgcggaggtt	12660
gctggcgaac	gcgacgacgc	ggcggttgat	ctcctgaatc	tggcgcctct	gcgtgaagac	12720
gacgggcccg	gtgagcttga	acctgaaaga	gagttcgaca	gaatcaattt	cggtgtcgtt	12780
gacggcggcc	tggcgcaaaa	tctcctgcac	gtctcctgag	ttgtcttgat	aggcgatctc	12840
				cgtccggctc		12900
ggcggcgagg	tcgttggaga	tgcgggccat	gagctgcgag	aaggcgttga	ggcctccctc	12960
gttccagacg	cggctgtaga	ccacgccccc	ttcggcatcg	cgggcgcgca	tgaccacctg	13020
cgcgagattg	agctccacgt	gccgggcgaa	gacggcgtag	tttcgcaggc	gctgaaagag	13080
gtagttgagg	gtggtggcgg	tgtgttctgc	cacgaagaag	tacataaccc	agcgccgcaa	13140
cgtggattcg	ttgatatccc	ccaaggcctc	aaggcgctcc	atggcctcgt	agaagtccac	13200
ggcgaagttg	aaaaactggg	agttgcgcgc	cgacacggtt	aactcctcct	ccagaagacg	13260
gatgageteg	gcgacagtgt	cgcgcacctc	gcgctcaaag	gctacagggg	cctcttcttc	13320
ttcttcaatc	tcctcttcca	taagggcctc	cccttcttct	tcttctggcg	gcggtggggg	13380
aggggggaca	cggcggcgac	gacggcgcac	cgggaggcgg	tcgacaaagc	gctcgatcat	13440
ctcccgcgg	cgacggcgca	tggtctcggt	gacggcgcgg	ccgttctcgc	gggggcgcag	13500
ttggaagacg	ccgcccgtca	tgtcccggtt	atgggttggc	ggggggctgc	cgtgcggcag	13560
ggatacggcg	ctaacgatgc	atctcaacaa	ttgttgtgta	ggtactccgc	caccgaggga	13620
				agaaaggcgt		13680
acagtcgcaa	ggtaggctga	gcaccgtggc	gggcggcagc	gggcggcggt	cggggttgtt	13740
tctggcggag	gtgctgctga	tgatgtaatt	aaagtaggcg	gtcttgagac	ggcggatggt	13800
cgacagaagc	accatgtcct	tgggtccggc	ctgctgaatg	cgcaggcggt	cggccatgcc	13860
ccaggcttcg	ttttgacatc	ggcgcaggtc	tttgtagtag	tcttgcatga	gcctttctac	13920
cggcacttct	tettetett	cctcttgtcc	tgcatctctt	gcatctatcg	ctgcggcggc	13980
ggcggagttt	ggccgtaggt	ggcgccctct	tcctcccatg	cgtgtgaccc	cgaagcccct	14040
catcggctga	agcagggcca	ggtcggcgac	aacgcgctcg	gctaatatgg	cctgctgcac	14100
ctgcgtgagg	gtagactgga	agtcgtccat	gtccacaaag	cggtggtatg	cgcccgtgtt	14160
gatggtgtaa	gtgcagttgg	ccataacgga	ccagttaacg	gtctggtgac	ccggctgcga	14220
gageteggtg	tacctgagac	gcgagtaagc	ccttgagtca	aagacgtagt	cgttgcaagt	14280
ccgcaccagg	tactggtatc	ccaccaaaaa	gtgcggcggc	ggctggcggt	agaggggcca	14340
gcgtagggtg	gccggggctc	cgggggcgag	gtcttccaac	ataaggcgat	gatatccgta	14400
gatgtacctg	gacatccagg	tgatgccggc	ggcggtggtg	gaggcgcgcg	gaaagtcacg	14460
gacgcggttc	cagatgttgc	gcagcggcaa	aaagtgctcc	atggtcggga	cgctctggcc	14520
ggtcaggcgc	gcgcagtcgt	tgacgctcta	gaccgtgcaa	aaggagagcc	tgtaagcggg	14580
cactcttccg	tggtctggtg	gataaattcg	caagggtatc	atggcggacg	accggggttc	14640
gaaccccgga	tccggccgtc	cgccgtgatc	catgcggtta	ccgcccgcgt	gtcgaaccca	14700
ggtgtgcgac	gtcagacaac	gggggagcgc	tccttttggc	ttccttccag	gcgcggcgga	14760
tgctgcgcta	gcttttttgg	ccactggccg	cgcgcggcgt	aagcggttag	gctggaaagc	14820
gaaagcatta	agtggctcgc	tccctgtagc	cggagggtta	ttttccaagg	gttgagtcgc	14880
gggacccccg	gttcgagtct	cgggccggcc	ggactgcggc	gaacgggggt	ttgcctcccc	14940
gtcatgcaag	accccgcttg	caaattcctc	cggaaacagg	gacgagcccc	ttttttgctt	15000
ttcccagatg	catccggtgc	tgcggcagat	gcgccccct	cctcagcagc	ggcaagagca	15060
agagcagcgg	cagacatgca	gggcaccctc	cccttctcct	accgcgtcag	gaggggcaac	15120
atccgcggct	gacgcggcgg	cagatggtga	ttacgaaccc	ccgcggcgcc	ggacccggca	15180
ctacttggac	ttggaggagg	gcgagggcct	ggcgcggcta	ggagcgccct	ctcctgagcg	15240
acacccaagg	gtgcagctga	agcgtgacac	gcgcgaggcg	tacgtgccgc	ggcagaacct	15300
gtttcgcgac	cgcgagggag	aggagcccga	ggagatgcgg	gatcgaaagt	tccatgcagg	15360

						15420
gcgcgagttg	cggcatggcc	tgaaccgcga	gcggttgctg	cgcgaggagg	actttgagcc	15420
caacacacaa	accoggatta	gtcccgcgcg	cgcacacgtg	gcggccgccg	acciggiaac	15480
cacatacaaa	cagacggtga	accaggagat	taactttcaa	aaaagcttta	acaaccacy	15540
gegeacgett	ataacacaca	aggaggtggc	tataggactg	atgcatctgt	gggactttgt	15600
aagcgcgctg	gagcaaaacc	caaatagcaa	gccgctcatg	gcgcagctgt	tecttatage	15660
gcagcacagc	agggacaacg	aggcattcag	ggatgcgctg	ctaaacatag	tagageeega	15720
agaccactag	ctoctcoatt	tgataaacat	.tctgcagagc	atagtggtgc	aggagegeag	15780
cttgagectg	actgacaagg	tggccgccat	taactattcc	atgctcagtc	tgggcaagtt	15840
ttacgcccgc	aagatatacc	atacccctta	cgttcccata	gacaaggagg	taaagatega	15900
ggggttctac	atococatoo	cgctgaaggt	gcttaccttg	agcgacgacc	tgggcgttta	15960
tracaacaaa	cgcatccaca	aggccgtgag	cgtgagccgg	cggcgcgagc	teagegaceg	16020
cgagctgatg	cacageetge	aaagggccct	ggctggcacg	ggcagcggcg	atagagaggc	16080
coantectae	tttgacgcgg	gcgctgacct	gcgctgggcc	ccaagccgac	gegeeetgya	16140
aacaactaaa	accadaccta	agctggcggt	ggcacccgcg	cgcgctggca	acgreggegg	16200
catagaggaa	tatgacgagg	acgatgagta	cgagccagag	gacggcgagt	actaagcygt	16260
gatgtttctg	atcagatgat	gcaagacgca	acggacccgg	cggtgcgggc	ggegetgeag	16320
agecagecgt	ccggccttaa	ctccacggac	gactggcgcc	aggtcatgga	ccgcaccarg	16380
tegetgactg	cacacaaccc	tgacgcgttc	cggcagcagc	cgcaggccaa	ceggetetee	16440
gcaattctgg	aagcggtggt	cccggcgcgc	gcaaacccca	cgcacgagaa	ggtgctggcg	16500
atcotaaaco	cactaaccaa	aaacagggcc	atccggcccg	atgaggccgg	cctggtctac	16560
gacgcgctgc	ttcagcgcgt	ggctcgttac	aacagcagca	acgtgcagac	caacctggac	16620
caactaataa	aggatataca	cgaggccgtg	gcgcagcgtg	agcgcgcgca	gcagcagggc	16680
aacctgggct	ccatggttgc	actaaacgcc	ttcctgagta	cacagcccgc	caacgtgccg	16740
caaaaacaaa	aggactacac	caactttgtg	agcgcactgc	ggctaatggt	gactgagaca	16800
ccacaaaata	aggtgtatca	gtccgggcca	gactattttt	tccagaccag	tagacaaggc	16860
ctgcagagcg	taaacctgag	ccaggettte	aagaacttgc	aggggctgtg	gggggtgcgg	16920
acteccacag	acaaccacac	gaccgtgtct	agcttgctga	cgcccaactc	gcgcctgttg	16980
ctactactaa	tagcgccctt	cacagacagt	ggcagcgtgt	cccgggacac	atacctaggt	17040
cogcogcoda	cactgtaccg	cgaggccata	agtcaggcgc	atgtggacga	gcatactttc	17100
caccegetga	caactottag	ccacacacta	gggcaggagg	acacgggcag	cctggaggca	17160
accetgaact	acctgctgac	caaccggcgg	caaaaaatcc	cctcgttgca	cagtttaaac	17220
acceegaace	agcgcatttt	acactatata	cagcagagcg	tgagccttaa	cctgatgcgc	17280
agegaggagg	cacceageat	gacactagac	atgaccgcgc	gcaacatgga	accgggcatg	17340
tateceteaa	acconcott	tatcaatcgc	ctaatggact	acttgcatcg	cgcggccgcc	17400
catgeeceaa	actigete	caatgccatc	ttgaacccgc	actggctacc	gcccctggt	17460
gryaaccccg	agaattaa	agtacccasa	ggtaacgatg	gattcctctg	ggacgacata	17520
cictacaccg	tatttcccc	acascacsa	accetoctao	agttgcaaca	acgcgagcag	17580
gacgacageg	cactacasas	gcaacogtag	cgcaggccaa	gcagettqte	cgatctaggc	17640
geagaggegg	cacaatcaaa	tactagtage	ccatttccaa	gcttgatagg	gtctcttacc	17700
getgeggeee	cacaccacca	acacctacta	gacgaggagg	agtacctaaa	caactcgctg	17760
agcactegea	accccacc	gaacctgcct	ccaacatttc	ccaacaacgg	gatagagagc	17820
ctgcagccgc	agegegataa	atogaagaco	tatgcgcagg	agcacaggga	tgtgcccggc	17880
ctagtggaca	agacgagtag	tcaaaggcag	gaccgtcage	aggatetagt	gtgggaggac	17940
cegegeeege	ccacccgccg	caccatctta	gatttgggag	ggagtggcaa	cccgtttgca	18000
gatgattegg	cayacyacay	cagegeees	taaaaaaaaa	catgatgcaa	aataaaaaac	18060
caccttegee	ccaggerggg	aggaatgett	trottatat	cccttagta	tgcggcgcgc	18120
tcaccaagge	catggcattg	agegregge	ctacqaqaqq	ataataaaca	cggcgccagt	18180
ggcgatgtat	gaggaaggcc	cettegatge	tecectagas	ccaccattca	tgcctccgcg	18240
ggcggcggcg	ctgggttcac	ccttcgatgc	. catccattac	totaaattaa	caccctatt	18300
gtacctgcgg	cetacegggg	yyayaaacay ttataacaa	, caectgetae	gatgtggcat	ccctgaacta	18360
cgacaccacc	cgtgtgtacc	ttotaaccac	· catcattcaa	aacaatgact	acagcccggg	18420
ccagaacgac	cacagcaact	toaatottoa	. ggtcactcae	cactggggg	gcgacctgaa	18480
ggaggcaagc	acacagacca	taccasatat	. cyaccygicy . caaccantto	atotygygcg	gcgacctgaa ataagtttaa	18540
aaccatcctg	cataccaaca	. cgccaaatgt	. yaacyayiii	. canatanaac	ataagtttaa tgaaatacga	18600
ggcgcgggtg	acggcgccgc	getegettat	. ctactcccac	. caggingage	tgaaatacga	18660
gtgggtggag	ttcacgctgc	ccgagggcaa	clacicogag	accargacca	tagaccttat	

gaacaacgcg	atcgtggagc	actacttgaa	agtgggcagg	cagaacgggg	ttctggaaag	18720
cgacatcggg	gtaaagtttg	acacccgcaa	cttcagactg	gggtttgacc	cagtcactgg	18780
tcttgtcatg	cctggggtat	atacaaacga	agccttccat	ccagacatca	ttttgctgcc	18840
					tccgcaagcg	18900
		ttaggatcac				18960
cgcactgttg	gatgtggacg	cctaccaggc	aagcttgaaa	gatgacaccg	aacagggcgg	19020
		acaacagtgg				19080
agctgcggca	atgcagccgg	tggaggacat	gaacgatcat	gccattcgcg	gcgacacctt	19140
tgccacacgg	gcggaggaga	agcgcgctga	ggccgaggca	gcggccgaag	ctgccgcccc	19200
cgctgcggag	gctgcacaac	ccgaggtcga	gaagcctcag	aagaaaccgg	tgattaaacc	19260
cctgacagag	gacagcaaga	aacgcagtta	caacctaata	agcaatgaca	gcaccttcac	19320
ccagtaccgc	agctggtacc	ttgcatacaa	ctacggcgac	cctcaggccg	ggatccgctc	19380
atggaccctg	ctttgcactc	ctgacgtaac	ctgcggctcg	gagcaggtat	actggtcgtt	19440
gcccgacatg	atgcaagacc	ccgtgacctt	ccgctccacg	cgccagatca	gcaactttcc	19500
ggtggtgggc	gccgagctgt	tgcccgtgca	ctccaagagc	ttctacaacg	accaggccgt	19560
ctactcccag	ctcatccgcc	agtttacctc	tctgacccac	gtgttcaatc	gctttcccga	19620
		cgccagcccc				19680
tgctctcaca	gatcacggga	cgctaccgct	gcgcaacagc	atcggaggag	tccagcgagt	19740
		gccgcacctg				19800
		gccgcacttt				19860
		gcctgcgctt				19920
		tgcgcgtgcg				19980
caaacgcggc	cgcactgggc	gcaccaccgt	cgatgacgcc	atcgacgcgg	tggtggagga	20040
ggcgcgcaac	tacacgccca	cgccgccgcc	agtgtccacc	gtggacgcgg	ccattcagac	20100
		gctacgctaa				20160
		gcactgccgc				20220
		gggcggccat				20280
		ccaggcgacg				20340
		ggggcaacgt				20400
		gcccccgcg				20460
		cagcggcggc				20520
		aggtcatcgc				20580
		gaaagctaaa				20640
		aggtggaact				20700
		taagacgtgt				20760
		gcacctacaa				20820
		acgagcgcct				20880
		tggacgaggg				20940
		cgcttgcacc				21000
		ccgtgcagct				21060
		tggagcctgg				21120
		tgggcgtgca				21180
tagcactagt	attoccacto	ccacagaggg	catqqaqaca	caaacqtccc	cggttgcctc	21240
		tgcaggcggc				21300
		tgtttcgtgt				21360
		cgctactgcc				21420
		acacctaccg				21480
		gccgccgtcg				21540
		aaggaggcag				21600
		cggtctttgt				21660
		gattccgagg				21720
		tgcgtcgtgc				21780
		tgcccctcct				21840
		tggccttgca				21900
		aaagtctgga				21960
Jacycygada		aaagtetgga		- 3 3 5	-3	

ttatanaata	gaagacatca	actttgcgtc	actooccco	cgacacggct	cgcgcccgtt	22020
catagaacg	tancaarata	tcggcaccag	caatatgagc	gatagcacct	tcagctgggg	22080
catgggaaac	agcaagata	aaaatttcgg	ttccaccatt	aagaactatg	gcagcaaagc	22140
ctcgctgtgg	ageggeacea	agatgctgag	ggacaagttg	aaagagcaaa	atttccaaca	22200
ctggaacage	gatggcctgg	cctctggcat	tagcagata	gtggacctgg	ccaaccaggc	22260
adaggiggia	aagattaaga	gtaagcttga	teceegeet	cccatagagg	agcctccacc	22320
agtgcaaaat	aayattaata	cagaggggcg	taacaaaaaa	catccacac	ccgacaggga	22380
ggccgtggag	atagegeeee	tagacgagcc	tccctcatac	gaggaggcac	taaagcaagg	22440
agaaactetg	gracycaaa	tcgcgcccat	aactaccaaa	atactagacc	agcacacacc	22500
cetgeedace	gacctgccca	ccccgccga	cacccaacaa	aaacctgtgc	taccagaccc	22560
cytaacycty	gacctgcccc	gtcctagccg	cacatacata	caccacacca	ccagcggtcc	22620
greegeegee	gergeaacee	ccagtggcaa	ctggcaaagc	acactgaaca	gcatcgtggg	22680
gegategitg	cygcccytag	agcgccgacg	atacttctaa	tagctaacgt	gtcgtatgtg	22740
trigggggrg	caatccctga	cgccgccaga	acgectecta	adccaccaca	cacccacttt	22800
tgtcatgtat	togacetta	atgatgccgc	agtgatgatg	catgcacatc	tcgggccagg	22860
ccaagatggc	cacccccccg	cccgggctgg	tacaattcac	ccacaccacc	gagacgtact	22920
acgcctcgga	gracergage	agaaacccca	caataacacc	tacqcacqac	gtgaccacag	22980
tcagcctgaa	taacaagttt	ctgcggttca	tececatana	ccacaaaaat	actocotact	23040
accggtctca	gegettgaeg	ctgcggttca	atastasca	tatactagac	atggcttcca	23100
cgtacaaggc	geggtteace	ctagctgtgg	graducetac	ttttaagccc	tactctggca	23160
cgtactttga	catecgegge	gtgctggaca	ggggcccac	atacaaataa	gaacaaaatg	23220
ctgcctacaa	egcactggcc	cccaagggtg	2222222	graceagrag	aatmaamctc	23280
aaactgcaca	agtggatgct	caagaacttg	acgaagagga	tacccaaact	ccactatcca	23340
aggcgcgaga	acaggaacaa	gctaagaaaa	taggatag	cascacases	ataccarata	23400
gaataaaaat	aactaaagaa	ggtctacaaa	taggaactgc	accacaacta	grageaggra	23460
ccggcaaaga	aattttcgca	gacaaaactt	ticaacciga	tattaaaaga	acaactccca	23520
aatggaacga	agcggatgcc	acagcagctg	gtggaagggt	tettaaaaay	acaaccccca	23580
tgaaaccctg	ctatggctca	tacgctagac	ccaccaattc	caacggcgga	tecacateca	23640
tggttgaaca	aaatggtaaa	ttggaaagtc	aagtegaaat	gcaattttt	cccacaccca	23700
caaatgccac	aaatgaagtt	aacaatatac	aaccaacagt	tgtattgtat	agcgaagatg	23760
taaacatgga	aactccagat	actcatcttt	cttataaacc	taaaatgggg	gataaaaaty	23700
ccaaagtcat	gcttggacaa	caagcaatgc	caaacagacc	aaattacatt	gettttagag	23880
acaattttat	tggtctcatg	tattacaaca	gcacaggtaa	catgggtgtc	ettgetggte	23940
aggcatcgca	gttgaacgct	gttgtagatt	tgcaagacag	aaacacagag	etgteetace	24000
agcttttgct	tgattcaatt	ggcgacagaa	caagatactt	ttcaatgtgg	aatcaagctg	24060
ttgacagcta	tgatccagat	gtcagaatta	ttgagaacca	tggaactgag	gatgagttgc	24120
caaattattg	ctttcctctt	ggtggaattg	ggattactga	cacttttcaa	gctgttaaaa	24120
caactgctgc	taacggggac	caaggcaata	ctacctggca	aaaagattca	acatttgcag	
aacgcaatga	aataggggtg	ggaaataact	ttgccatgga	aattaacctg	aatgccaacc	24240
tatggagaaa	tttcctttac	tccaatattg	cgctgtacct	gccagacaag	ctaaaataca	24300 24360
accccaccaa	tgtggaaata	tctgacaacc	ccaacaccta	cgactacatg	aacaagcgag	
tggtggctcc	tgggcttgta	gactgctaca	ttaaccttgg	ggcgcgctgg	tctctggact	24420
acatggacaa	cgttaatccc	tttaaccacc	accgcaatgc	gggcctgcgt	taccgctcca	24480
tgttgttggg	aaacggccgc	tacgtgccct	ttcacattca	ggtgcccaa	aagtttttg	24540
ccattaaaaa	cctcctcctc	ctgccaggct	catacacata	tgaatggaac	ttcaggaagg	24600
atottaacat	ggttctgcag	agctctctgg	gaaacgacct	tagagttgac	ggggctagca	24660
ttaagtttga	cagcatttgt	ctttacgcca	ccttcttccc	catggcccac	aacacggcct	24720
ccacactaga	agccatgctc	agaaatgaca	ccaacgacca	gtcctttaat	gactaccttt	24780
ccaccaccaa	catgctatat	cccatacccg	ccaacgccac	caacgtgccc	atctccatcc	24840
catcgcgcaa	ctgggcagca	tttcgcggtt	gggccttcac	acgcttgaag	acaaaggaaa	24900
ccccttccct	gggatcaggc	tacgaccctt	actacaccta	ctctggctcc	ataccatacc	24960
ttgacggaac	cttctatctt	aatcacacct	ttaagaaggt	ggccattact	tttgactctt	25020
ctattageta	gccgggcaac	gaccgcctgc	ttactcccaa	tgagtttgag	attaagcgct	25080
cagttgacgg	ggagggctat	aacgtagctc	agtgcaacat	gacaaaggac	tggttcctag	25140
tocagatott	ggccaactac	aatattggct	accagggctt	ctacattcca	gaaagctaca	25200
aagaccgcat	gtactcqttc	ttcagaaact	tccagcccat	gagccggcaa	gtggtggacg	25260
		_				

atactaaata	caaagattat	cagcaggttg	gaattatcca	ccagcataac	aactcaggct	25320
tcgtaggcta	cctcgctccc	accatgcgcg	agggacaagc	ttaccccgct	aatgttccct	25380
acccactaat	aggcaaaacc	gcggttgata	gtattaccca	gaaaaagttt	ctttgcgacc	25440
gcaccctgtg	gcgcatcccc	ttctccagta	actttatgtc	catgggtgcg	ctcacagacc	25500
					tttgaggtgg	25560
atcccatgga	cgagcccacc	cttctttatg	ttttgtttga	agtctttgac	gtggtccgtg	25620
					ttctcggccg	25680
					tgggctccag	25740
					ttttgggcac	25800
					ccatagttaa	25860
cacggccggt	cgcgagactg	ggggcgtaca	ctggatggcc	tttgcctgga	acccgcgctc	25920
					agcaggttta	25980
					ccgaccgctg	26040
tataacgctg	gaaaagtcca	cccaaagcgt	gcaggggccc	aactcggccg	cctgtggcct	26100
attctgctgc	atgtttctcc	acgcctttgc	caactggccc	caaactccca	tggatcacaa	26160
					cccaggtaca	26220
					actcgcccta	26280
	cacagtgcgc					26340
	tgtactagga					26400
tcgggtgatt	atttaccccc	acccttgccg	tctgcgccgt	ttaaaaatca	aaggggttct	26460
	gctatgcgcc					26520
acttaaactc	aggcacaacc	atccgcggca	gctcggtgaa	gttttcactc	cacaggctgc	26580
gcaccatcac	caacgcgttt	agcaggtcgg	gcgccgatat	cttgaagtcg	cagttggggc	26640
ctccgccctg	cgcgcgcgag	ttgcgataca	cagggttaca	gcactggaac	actatcagcg	26700
	cacgctggcc					26760
	cagggcgaac					26820
	tgagttgcac					26880
	atacagcgcc					26940
	agagaagaac					27000
	cacgcagcac					27060
	cacgatcttg					27120
	atccatttca					27180
	gccttcgatc					27240
cgtggtgctt	gtaggttacc	tctgcaaacg	actgcaggta	cgcctgcagg	aatcgcccca	27300
	aaaggtcttg					27360
	cttgcatacg					27420
	atcgttatcc					27480
	cgcagacacg					27540
	ggactcttcc					27600
cttcattcag	ccgccgcacc	gtgcgcttac	ctcccttgcc	gtgcttgatt	agcaccggtg	27660
	acccaccatt					27720
	ggatggcggg					27780
	caaatccgcc					27840
	tgacgagtct					27900
	gcggggaggc					27960
	cgccgcaccg					28020
	ttccttctcc					28080
	cgccccttt					28140
	ccccgtcgag					28200
	tgtaagcgaa					28260
	cgacgcagag					28320
	agatgtggga					28380
	cgcgttgcaa					28440
	acgccacctg					28500
catgcgagcc	caacccgcgc	ctcaacttct	accccgtatt	tgccgtgcca	gaggtgcttg	28560

		casasctoca	agatacccct	atcctgccgt	gccaaccgca	28620
ccacctatca	catcuttuc	caaaactyca	agggcgctgt	catacctgat	atcocctcgc	28680
gccgagcgga	caagcagctg	tttaagate	ttggacgcga	cgagaagcgC	gcggcaaacg	28740
tcgacgaagt	gccaaaaacc	trans	gtcactgtgg	agtactagta	gaacttgagg	28800
ctctgcaaca	agaaaacagc	gaaaacgaaa	gceactgcgg	agtgaccac	tttacctacc	28860
gtgacaacgc	gegeetagee	gractiatas	gcagcatcga	gagcgagctg	atcatacacc	28920
cggcacttaa	cctaccccc	aaggccacga	gcacagtcat	aaccgaggag	ggcctacccg	28980
gtgcacgacc	cctggagagg	gatgeaaact	tgcaagaaca ttgagacgcg	cgagcctgcc	gacttggagg	29040
cagttggcga	tgagcagctg	gegegeegge	ttgttaccgt	gagettgag	tocatocage	29100
agcgacgcaa	gctaatgatg	geegeagege	agctagagga	aacottocac	tacacctttc	29160
ggttctttgc	tgacccggag	acgcagcgca	tttccaacat	ggagetetge	aacctggtct	29220
gccagggcta	cgtgcgccag	gcccycaaaa	tttccaacgt	catacttcat	tecaceetca	29280
cctaccttgg	aattttgcac	gaaaaccycc	ttgggcaaaa	cttatttctg	toctacacct	29340
agggcgaggc	gegeegegae	tacgcccgcg	actgcgttta	gcgcaacctg	aaggagctgc	29400
ggcaaacggc	catgggcgtg	tggcaycagc	gcctggagga	cttcaacgag	cactccataa	29460
agaagctgct	aaagcaaaac	ttgaaggacc	tatggacggc	taaaaccctq	caacagggtc	29520
ccgcgcacct	ggcggacatt	acceccec	aacgcctgct	gaactttatc	ctagagcgtt	29580
tgccagactt	caccagtcaa	agcatgttyc	aaaactttag	ctttataccc	attaagtacc	29640
caggaattct	gcccgccacc	tgetgtgege	ttcctagcga	gragetagee	aactaccttg	29700
gtgaatgccc	teegeegett	tggggtcact	gccaccccc	cctactogag	aactaccttg	29760
cctaccactc	cgacatcatg	gaagacgtga	teeteteeaa	ttcacaactq	tgtcactgtc cttagcgaaa	29820
gctgcaacct	atgcaccccg	caccgctccc	ctcctccc	taacaaaaa	cttagcgaaa	29880
gtcaaattat	cggtaccttt	gagetgeagg	gccccccgcc	ccttcccaaa	tccgcggctc	29940
cggggttgaa	actcactccg	gggctgtgga	egeeggeeea	atecegeaaa	tttgtacctg	30000
aggactacca	cgcccacgag	attaggttet	acyaayacca	cceattacaa	ccaaatgcgg	30060
agcttaccgc	ctgcgtcatt	acccagggcc	acatectegg	ttacttgcac	gccattaaca	30120
aagcccgcca	agagtttctg	ctacgaaagg	gacggggggt	teacceggae	ccccagtccg	30180
gcgaggagct	caacccaatc	ccccgccgc	egeageeeta	cagcagecg	cgggcccttg	30240
cttcccagga	tggcacccaa	aaagaagctg	cagetyeege	aggaggagg	cacggacgag	30300
gaggaatact	gggacagtca	ggcagaggag	gttttggatg	aggaggagga	gatgatggaa	30360
gactgggaca	gcctagacga	ggaagcttcc	gaggeegaay	coccaaccot	cgaaacaccg	30420
tcaccctcgg	tegeattece	ctcgccggcg	ccccayaaac	ttccccacc	tcccagcatt	30480
gctacaacct	ccgctcctca	ggcgccgccg	tetaageeeg	caccaccatt	caaccgtaga	30540
tgggacacca	ctggaaccag	ggccggtaay	cccaagcagc	ageagacat	agcccaagag	30600
caacaacagc	gccaaggcta	ccgctcgtgg	egegtgeaca	ttottotota	agttgcttgc	30660
ttgcaagact	gtgggggcaa	catctccttc	geeegeegee	tatagaggg	ccatcacggc	30720
gtggccttcc	cccgtaacat	cctgcattac	taccgtcatc	nagagagaga	ctactgcacc	30780
ggcggcagcg	gcagcaacag	cagcggccac	gcagaagcaa	aggegacegg	atagcaagac	30840
tctgacaaag	cccaagaaat	ccacagcggc	ggcagcagca	. ggaggaggag	cactgcgtct	30900
ggcgcccaac	gaacccgtat	cgacccgcga	gcttagaaac	ayyaccccc	ccactctgta	30960
tgctatattt	caacagagca	ggggccaaga	acaagagctg	aaaacaaaaa	acaggtetet	31020
gcgctccctc	acccgcagct	gcctgtatca	caaaagcyaa	gattagette	ggcgcacgct	31080
ggaagacgcg	gaggetetet	tcagcaaata	ctgegegete	caccacaca	actagtttcg	31140
cgccctttct	caaatttaag	cgcgaaaact	acguatece	. cageggeeae	acceggegee	31200
accacctoto	atcadodoca	l ttatgagcaa	. ggaaattici	, acyceetaet	. cgcggmgcom	31260
ccagccacaa	atgggacttg	cggctggagc	tgcccaayac	- tactcaaccc	gaataaacta	31320
catgagcgcg	ggaccccaca	tgatatcccg	ggtcaacgga	accegegeee	accgaaaccg	31380
aattctcctc	gaacaggcgg	, ctattaccac	cacacctcgt	aataacctta	atccccgtag	31440
ttggcccgct	gccctggtgt:	: accaggaaag	tecegeteed	accactgigg	tacttcccag	31500
agacgcccag	gccgaagtto	agatgactaa	ctcaggggcg	cagettgegg	geggettteg	31560
teacagggt	reauteacec	r aacaaaatat	: aactcacctg	j aaaatcagag	gycyaggiai	31620
+ a a g c t c a a c	· dacdadtcdd	, toagctcctc	: tcttggtctc	: cgrccggacg	gyacacccca	31680
	actaaccact	- cttcatttac	: qccccgtcag	y gcgattctta	Cicigcagac	31740
atactactac	, gageededet	ccagaggcat	tggaactcta	a caatttattg	aygagicege	31800
atteagtt	 tacttcaace 	e cattttata	acctcccgg	cactacecy:	accageeeae	
tcccaacttt	gacgcggtaa	a aagactcgg	ggacggctad	: gactgaatga	a ccagtggaga	31000

ggcagagcaa	ctgcgcctga	cacacctcga	ccactgccgc	cgccacaagt	gctttgcccg	31920
cggctccggt	gagttttgtt	actttgaatt	gcccgaagag	catatcgagg	gcccggcgca	31980
cggcgtccgg	ctcaccaccc	aggtagagct	tacacgtagc	ctgattcggg	agtttaccaa	32040
gcgccccctg	ctagtggagc	gggagcgggg	tccctgtgtt	ctgaccgtgg	tttgcaactg	32100
tcctaaccct	ggattacatc	aagatcttat	tccattcaac	taacaataaa	cacacaataa	32160
attacttact	taaaatcagt	cagcaaatct	ttgtccagct	tattcagcat	cacctccttt	32220
ccctcctccc	aactctggta	tttcagcagc	cttttagctg	cgaactttct	ccaaagtcta	. 32280
aatgggatgt	caaattcctc	atgttcttgt	ccctccgcac	ccactatctt	catattgttg	32340
cagatgaaac	gcgccagacc	gtctgaagac	accttcaacc	ctgtgtaccc	atatgacacg	32400
gaaaccggcc	ctccaactgt	gcctttcctt	acccctccct	ttgtgtcgcc	aaatgggttc	32460
caagaaagtc	ccccggagt	gctttctttg	cgtctttcag	aacctttggt	tacctcacac	32520
ggcatgcttg	cgctaaaaat	gggcagcggc	ctgtccctgg	atcaggcagg	caaccttaca	32580
tcaaatacaa	tcactgtttc	tcaaccgcta	aaaaaaacaa	agtccaatat	aactttggaa	32640
acatccgcgc	cccttacagt	cagctcaggc	gccctaacca	tggccacaac	ttcgcctttg	32700
gtggtctctg	acaacactct	taccatgcaa	tcacaagcac	cgctaaccgt	gcaagactca	32760
aaacttagca	ttgctaccaa	agagccactt	acagtgttag	atggaaaact	ggccctgcag	32820
acatcagccc	ccctctctgc	cactgataac	aacgccctca	ctatcactgc	ctcacctcct	32880
cttactactg	caaatggtag	tctggctgtt	accatggaaa	acccacttta	caacaacaat	32940
ggaaaacttg	ggctcaaaat	tggcggtcct	ttgcaagtgg	ccaccgactc	acatgcacta	33000
acactaggta	ctggtcaggg	ggttgcagtt	cataacaatt	tgctacatac	aaaagttaca	33060
ggcgcaatag	ggtttgatac	atctggcaac	atggaactta	aaactggaga	tggcctctat	33120
gtggatagcg	ccggtcctaa	ccaaaaacta	catattaatc	taaataccac	aaaaggcctt	33180
gcttttgaca	acaccgcaat	aacaattaac	gctggaaaag	ggttggaatt	tgaaacagac	33240
tcctcaaacg	gaaatcccat	aaaaacaaaa	attggatcag	gcatacaata	taataccaat	33300
ggagctatgg	ttgcaaaact	tggaacaggc	ctcagttttg	acagctccgg	agccataaca	33360
	taaacaatga					33420
tgcagaattg	cttcagataa	agactgcaag	ctaactctgg	cgctaacaaa	atgtggcagt	33480
caaattttgg	gcactgtttc	agctttggca	gtatcaggta	atatggcctc	catcaatgga	33540
actctaagca	gtgtaaactt	ggttcttaga	tttgatgaca	acggagtgct	tatgtcaaat	33600
tcatcactgg	acaaacagta	ttggaacttt	agaaacgggg	actccactaa	cggtcaacca	33660
tacacttatg	ctgttgggtt	tatgccaaac	ctaaaagctt	acccaaaaac	tcaaagtaaa	33720
actgcaaaaa	gtaatattgt	tagccaggtg	tatcttaatg	gtgacaagtc	taaaccattg	33780
cattttacta	ttacgctaaa	tggaacagat	gaaaccaacc	aagtaagcaa	atactcaata	33840
tcattcagtt	ggtcctggaa	cagtggacaa	tacactaatg	acaaatttgc	caccaattcc	33900
tataccttct	cctacattgc	ccaggaataa	agaatcgtga	acctgttgca	tgttatgttt	33960
caacgtgttt	atttttcaat	tgcagaaaat	ttcaagtcat	ttttcattca	gtagtatagc	34020
cccaccacca	catagcttat	actaatcacc	gtaccttaat	caaactcaca	gaaccctagt	34080
attcaacctg	ccacctccct	cccaacacac	agagtacaca	gtcctttctc	cccggctggc	34140
cttaaacagc	atcatatcat	gggtaacaga	catattctta	ggtgttatat	tccacacggt	34200
ctcctgtcga	gccaaacgct	catcagtgat	gttaataaac	tccccgggca	gctcgcttaa	34260
gttcatgtcg	ctgtccagct	gctgagccac	aggctgctgt	ccaacttgcg	gttgctcaac	34320
gggcggcgaa	ggagaagtcc	acgcctacat	gggggtagag	tcataatcgt	gcatcaggat	34380
agggcggtgg	tgctgcagca	gcgcgcgaat	aaactgctgc	cgccgccgct	ccgtcctgca	34440
ggaatacaac	atggcagtgg	tctcctcagc	gatgattcgc	accgcccgca	gcataaggcg	34500
ccttgtcctc	cgggcacagc	agcgcaccct	gatctcactt	aagtcagcac	agtaactgca	34560
	acaatattgt					34620
	acagaaccca					34680
	aacacgctgg					34740
ctcccggtac	catataaacc	tctgattaaa	catggcgcca	tccaccacca	tcctaaacca	34800
gctggccaaa	acctgcccgc	cggctatgca	ctgcagggaa	ccgggactgg	aacaatgaca	34860
gtggagagcc	caggactcgt	aaccatggat	catcatgctc	gtcatgatat	caatgttggc	34920
acaacacagg	cacacgtgca	tacacttcct	caggattaca	agctcctccc	gcgtcagaac	34980
catatcccag	ggaacaaccc	attcctgaat	cagcgtaaat	cccacactgc	agggaagacc	35040
tcgcacgtaa	ctcacgttgt	gcattgtcaa	agtgttacat	tcgggcagca	gcggatgatc	35100
ctccagtatg	gtagcgcggg	tttctgtctc	aaaaggaggt	agacgatccc	tactgtacgg	35160

```
agtgcgccga gacaaccgag atcgtgttgg tcgtagtgtc atgccaaatg gaacgccgga
cgtagtcata tttcctgaag caaaaccagg tgcgggcgtg acaaacagat ctgcgtctcc
ggtctcgccg cttagatcgc tctgtgtagt agttgtagta tatccactct ctcaaagcat
                                                                  35340
ccaggegece cetggetteg ggttetatgt aaacteette atgegeeget geeetgataa
                                                                  35400
catccaccac cgcagaataa gccacaccca gccaacctac acattcgttc tgcgagtcac
                                                                  35460
acacgggagg agcgggaaga gctggaagaa ccatgttttt ttttttattc caaaagatta
                                                                  35520
tecaaaacet caaaatgaag atetattaag tgaacgeget ecceteeggt ggegtggtea
                                                                  35580
aactctacag ccaaagaaca gataatggca tttgtaagat gttgcacaat ggcttccaaa
                                                                  35640
aggcaaacgg ccctcacgtc caagtggacg taaaggctaa acccttcagg gtgaatctcc
                                                                  35700
totataaaca ttocagcaco ttoaaccatg cocaaataat totoatotog coaccttoto
                                                                  35760
aatatatoto taagoaaato oogaatatta agtooggoca ttgtaaaaat otgotooaga
                                                                  35820
gegeceteca cetteageet caageagega atcatgattg caaaaattca ggtteeteac
agacctgtat aagattcaaa agcggaacat taacaaaaat accgcgatcc cgtaggtccc
ttcgcagggc cagctgaaca taatcgtgca ggtctgcacg gaccagcgcg gccacttccc
                                                                  36000
cgccaggaac catgacaaaa gaacccacac tgattatgac acgcatactc ggagctatgc
                                                                  36060
taaccagcgt agccccgatg taagcttgtt gcatgggcgg cgatataaaa tgcaaggtgc
                                                                  36120
tgctcaaaaa atcaggcaaa gcctcgcgca aaaaagaaag cacatcgtag tcatgctcat
                                                                  36180
gcagataaag gcaggtaagc tccggaacca ccacagaaaa agacaccatt tttctctcaa
                                                                  36240
36300
agaagcctgt cttacaacag gaaaaacaac ccttataagc ataagacgga ctacggccat
                                                                  36360
gccggcgtga ccgtaaaaaa actggtcacc gtgattaaaa agcaccaccg acagctcctc
                                                                  36420
ggtcatgtcc ggagtcataa tgtaagactc ggtaaacaca tcaggttgat tcacatcggt
                                                                  36480
cagtgctaaa aagcgaccga aatagcccgg gggaatacat acccgcaggc gtagagacaa
                                                                  36540
cattacagcc cccataggag gtataacaaa attaatagga gagaaaaaca cataaacacc
                                                                  36600
tgaaaaaccc tcctgcctag gcaaaatagc accctcccgc tccagaacaa catacagcgc
                                                                  36660
ttccacagcg gcagccataa cagtcagcct taccagtaaa aaagaaaacc tattaaaaaa
                                                                  36720
acaccactcg acacggcacc agetcaatca gtcacagtgt aaaaaaggge caagtgcaga
                                                                  36780
gcgagtatat ataggactaa aaaatgacgt aacggttaaa gtccacaaaa aacacccaga
                                                                  36840
                                                                  36900
aaaccgcacg cgaacctacg cccagaaacg aaagccaaaa aacccacaac ttcctcaaat
cgtcacttcc gttttcccac gttacgtcac ttcccatttt aagaaaacta caattcccaa
                                                                  36960
cacatacaag ttactccgcc ctaaaaccta cgtcacccgc cccgttccca cgccccgcgc
                                                                   37020
cacgtcacaa actccaccc ctcattatca tattggcttc aatccaaaat aaggtatatt
                                                                   37080
                                                                   37090
attgatgatg
<210> 5
<211> 5955
<212> DNA
<213> Artificial Sequence
<220>
<223> NS cDNA sequence
<221> CDS
<222> (1)...(5955)
<400> 5
atg gcg ccc atc acg gcc tac tcc caa cag acg cgg ggc cta ctt ggt
                                                                     48
Met Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly
                 5
 1
                                                                     96
tgc atc atc act agc ctt aca ggc cgg gac aag aac cag gtc gag gga
Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
             20
gag gtt cag gtg gtt tcc acc gca aca caa tcc ttc ctg gcg acc tgc
                                                                    144
```

Glu	Val	Gln 35	Val	Val	Ser	Thr	Ala 40	Thr	Gln	Ser	Phe	Leu 45	Ala	Thr	Cys	٠.
gtc Val	aac Asn 50	ggc Gly	gtg Val	tgt Cys	tgg Trp	acc Thr 55	gtt Val	tac Tyr	cat His	ggt Gly	gct Ala 60	ggc Gly	tca Ser	aag Lys	acc Thr	192
tta Leu 65	gcc Ala	ggc Gly	cca Pro	aag Lys	ggg Gly 70	cca Pro	atc Ile	acc Thr	cag Gln	atg Met 75	tac Tyr	act Thr	aat Asn	gtg Val	gac Asp 80	240
cag Gln	gac Asp	ctc Leu	gtc Val	ggc Gly 85	tgg Trp	cag Gln	gcg Ala	ccc Pro	ccc Pro 90	Gly ggg	gcg Ala	cgt Arg	tcc Ser	ttg Leu 95	aca Thr	288
cca Pro	tgc Cys	acc Thr	tgt Cys 100	ggc Gly	agc Ser	tca Ser	gac Asp	ctt Leu 105	tac Tyr	ttg Leu	gtc Val	acg Thr	aga Arg 110	cat His	gct Ala	336
gac Asp	gtc Val	att Ile 115	ccg Pro	gtg Val	cgc Arg	cgg Arg	cgg Arg 120	ggc Gly	gac Asp	agt Ser	agg Arg	ggg Gly 125	agc Ser	ctg Leu	ctc Leu	384
tcc Ser	ccc Pro 130	agg Arg	cct Pro	gtc Val	tcc Ser	tac Tyr 135	ttg Leu	aag Lys	ggc Gly	tct Ser	tcg Ser 140	ggt Gly	ggt Gly	cca Pro	ctg Leu	432
ctc Leu 145	tgc Cys	cct Pro	tcg Ser	Gly	cac His 150	gct Ala	gtg Val	ggc Gly	atc Ile	ttc Phe 155	cgg Arg	gct Ala	gcc Ala	gta Val	tgc Cys 160	480
acc Thr	cgg Arg	Gly	gtt Val	gcg Ala 165	aag Lys	gcg Ala	gtg Val	gac Asp	ttt Phe 170	gtg Val	ccc Pro	gta Val	gag Glu	tcc Ser 175	atg Met	528
gaa Glu	act Thr	act Thr	atg Met 180	cgg Arg	tct Ser	ccg Pro	gtc Val	ttc Phe 185	acg Thr	gac Asp	aac Asn	tca Ser	tcc Ser 190	ccc Pro	ccg Pro	576
gcc Ala	gta Val	ccg Pro 195	cag Gln	tca Ser	ttt Phe	caa Gln	gtg Val 200	gcc Ala	cac His	cta Leu	cac His	gct Ala 205	ccc Pro	act Thr	ggc Gly	624
agc Ser	ggc Gly 210	aag Lys	agt Ser	act Thr	aaa Lys	gtg Val 215	ccg Pro	gct Ala	gca Ala	tat Tyr	gca Ala 220	gcc Ala	caa Gln	Gly ggg	tac Tyr	672
aag Lys 225	gtg Val	ctc Leu	gtc Val	ctc Leu	aat Asn 230	ccg Pro	tcc Ser	gtt Val	gcc Ala	gct Ala 235	acc Thr	tta Leu	GJÀ aaa	ttt Phe	ggg Gly 240	720
gcg Ala	tat Tyr	atg Met	tct Ser	aag Lys 245	gca Ala	cac His	ggt Gly	att Ile	gac Asp 250	ccc Pro	aac Asn	atc Ile	aga Arg	act Thr 255	GJÀ ààà	768

gta Val	agg Arg	acc Thr	att Ile 260	acc Thr	aca Thr	ggc	gcc Ala	ccc Pro 265	gtc Val	aca Thr	tac Tyr	tct Ser	acc Thr 270	tat Tyr	ggc Gly	816
aag Lys	ttt Phe	ctt Leu 275	gcc Ala	gat Asp	ggt Gly	ggt Gly	tgc Cys 280	tct Ser	ggg Gly	ggc Gly	gct Ala	tat Tyr 285	gac Asp	atc Ile	ata Ile	864
ata Ile	tgt Cys 290	gat Asp	gag Glu	tgc Cys	cat His	tca Ser 295	act Thr	gac Asp	tcg Ser	act Thr	aca Thr 300	atc Ile	ttg Leu	ggc Gly	atc Ile	912
ggc Gly 305	aca Thr	gtc Val	ctg Leu	gac Asp	caa Gln 310	gcg Ala	gag Glu	acg Thr	gct Ala	gga Gly 315	gcg Ala	cgg Arg	ctt Leu	gtc Val	gtg Val 320	960
ctc Leu	gcc Ala	acc Thr	gct Ala	acg Thr 325	cct Pro	ccg Pro	gga Gly	tcg Ser	gtc Val 330	acc Thr	gtg Val	cca Pro	cac His	cca Pro 335	aac Asn	1008
atc Ile	gag Glu	gag Glu	gtg Val 340	gcc Ala	ctg Leu	tct Ser	aat Asn	act Thr 345	gga Gly	gag Glu	atc Ile	ccc Pro	ttc Phe 350	tat Tyr	ggc Gly	1056
aaa Lys	gcc Ala	atc Ile 355	ccc Pro	att Ile	gaa Glu	gcc Ala	atc Ile 360	agg Arg	GJA aaa	gga Gly	agg Arg	cat His 365	ctc Leu	att Ile	ttc Phe	1104
tgt Cys	cat His 370	tcc Ser	aag Lys	aag Lys	aag Lys	tgc Cys 375	gac Asp	gag Glu	ctc Leu	gcc Ala	gca Ala 380	aag Lys	ctg Leu	tca Ser	ggc Gly	1152
ctc Leu 385	gga Gly	atc Ile	aac Asn	gct Ala	gtg Val 390	gcg Ala	tat Tyr	tac Tyr	cgg Arg	ggg Gly 395	ctc Leu	gat Asp	gtg Val	tcc Ser	gtc Val 400	1200
ata Ile	cca Pro	act Thr	atc Ile	gga Gly 405	gac Asp	gtc Val	gtt Val	gtc Val	gtg Val 410	gca Ala	aca Thr	gac Asp	gct Ala	ctg Leu 415	atg Met	1248
acg Thr	ggc Gly	tat Tyr	acg Thr 420	ggc Gly	gac Asp	ttt Phe	gac Asp	tca Ser 425	gtg Val	atc Ile	gac Asp	tgt Cys	aac Asn 430	aca Thr	tgt Cys	1296
gtc Val	acc Thr	cag Gln 435	aca Thr	gtc Val	gac Asp	ttc Phe	agc Ser 440	ttg Leu	gat Asp	ccc Pro	acc Thr	ttc Phe 445	acc Thr	att Ile	gag Glu	1344
acg Thr	acg Thr 450	acc Thr	gtg Val	cct Pro	caa Gln	gac Asp 455	gca Ala	gtg Val	tcg Ser	cgc Arg	tcg Ser 460	cag Gln	cgg Arg	cgg Arg	ggt Gly	1392
agg Arg 465	act Thr	ggc Gly	agg Arg	ggt Gly	agg Arg 470	aga Arg	ggc	atc Ile	tac Tyr	agg Arg 475	Phe	gtg Val	act Thr	ccg Pro	gga Gly 480	1440

					atg Met											1488
gac Asp	gcg Ala	ggc	tgt Cys 500	gct Ala	tgg Trp	tac Tyr	gag Glu	ctc Leu 505	acc Thr	ccc Pro	gcc Ala	gag Glu	acc Thr 510	tcg Ser	gtt Val	1536
					ctg Leu											1584
cac His	ctg Leu 530	gag Glu	ttc Phe	tgg Trp	gag Glu	agt Ser 535	gtc Val	ttc Phe	aca Thr	ggc Gly	ctc Leu 540	acc Thr	cac His	ata Ile	gat Asp	1632
gca Ala 545	cac His	ttc Phe	ttg Leu	tcc Ser	cag Gln 550	acc Thr	aag Lys	cag Gln	gca Ala	gga Gly 555	gac Asp	aac Asn	ttc Phe	ccc Pro	tac Tyr 560	1680
ctg Leu	gta Val	gca Ala	tac Tyr	caa Gln 565	gcc Ala	acg Thr	gtg Val	tgc Cys	gcc Ala 570	agg Arg	gct Ala	cag Gln	gcc Ala	cca Pro 575	cct Pro	1728
cca Pro	tca Ser	tgg Trp	gat Asp 580	caa Gln	atg Met	tgg Trp	aag Lys	tgt Cys 585	ctc Le u	ata Ile	cgg Arg	ctg Leu	aaa Lys 590	cct Pro	acg Thr	1776
					ccc Pro											1824
gag Glu	gtc Val 610	acc Thr	ctc Leu	acc Thr	cac His	ccc Pro 615	ata Ile	acc Thr	aaa Lys	tac Tyr	atc Ile 620	atg Met	gca Ala	tgc Cys	atg Met	1872
tcg Ser 625	Ala	gac Asp	ctg Leu	gag Glu	gtc Val 630	gtc Val	act Thr	agc Ser	acc Thr	tgg Trp 635	gtg Val	ctg Leu	gtg Val	ggc Gly	gga Gly 640	1920
gtc Val	ctt Leu	gca Ala	gct Ala	ctg Leu 645	gcc Ala	gcg Ala	tat Tyr	tgc Cys	ctg Leu 650	aca Thr	aca Thr	ggc Gly	agt Ser	gtg Val 655	gtc Val	1968
					atc Ile											2016
					cag Gln											2064
cac	ctc	cct	tac	atc	gag	cag	gga	atg	cag	ctc	gcc	gag	caa	ttc	aag	2112

His	Leu 690	Pro	Tyr	Ile	Glu	Gln 695	Gly	Met	Gln	Leu	Ala 700	Glu	Gln	Phe	Lys	
cag Gln 705	aaa Lys	gcg Ala	ctc Leu	ggg	tta Leu 710	ctg Leu	caa Gln	aca Thr	gcc Ala	acc Thr 715	aaa Lys	caa Gln	gcg Ala	gag Glu	gct Ala 720	2160
gct Ala	gct Ala	ccc Pro	gtg Val	gtg Val 725	gag Glu	tcc Ser	aag Lys	tgg Trp	cga Arg 730	gcc Ala	ctt Leu	gag Glu	aca Thr	ttc Phe 735	tgg Trp	2208
gcg Ala	aag Lys	cac His	atg Met 740	tgg Trp	aat Asn	ttc Phe	atc Ile	agc Ser 745	Gly ggg	ata Ile	cag Gln	tac Tyr	tta Leu 750	gca Ala	ggc Gly	2256
tta Leu	tcc Ser	act Thr 755	ctg Leu	cct Pro	ggg ggg	aac Asn	ccc Pro 760	gca Ala	ata Ile	gca Ala	tca Ser	ttg Leu 765	atg Met	gca Ala	ttc Phe	2304
aca Thr	gcc Ala 770	tct Ser	atc Ile	acc Thr	agc Ser	ccg Pro 775	ctc Leu	acc Thr	acc Thr	caa Gln	agt Ser 780	acc Thr	ctc Leu	ctg Leu	ttt Phe	2352
aac Asn 785	atc Ile	ttg Leu	Gly aaa	ggg Gly	tgg Trp 790	gtg Val	gct Ala	gcc Ala	caa Gln	ctc Leu 795	gcc Ala	ccc Pro	ccc Pro	agc Ser	gcc Ala 800	2400
gct Ala	tcg Ser	gct Ala	ttc Phe	gtg Val 805	ggc Gly	gcc Ala	ggc Gly	atc Ile	gcc Ala 810	ggt Gly	gcg Ala	gct Ala	gtt Val	ggc Gly 815	agc Ser	2448
ata Ile	ggc Gly	ctt Leu	ggg Gly 820	aag Lys	gtg Val	ctt Leu	gtg Val	gac Asp 825	att Ile	ctg Leu	gcg Ala	ggt Gly	tat Tyr 830	gga Gly	gca Ala	2496
gga Gly	gtg Val	gcc Ala 835	ggc Gly	gcg Ala	ctc Leu	gtg Val	gcc Ala 840	ttc Phe	aag Lys	gtc Val	atg Met	agc Ser 845	ggc Gly	gag Glu	atg Met	2544
ccc Pro	tcc Ser 850	acc Thr	gag Glu	gac Asp	ctg Leu	gtc Val 855	aat Asn	cta Leu	ctt Leu	cct Pro	gcc Ala 860	atc Ile	ctc Leu	tct Ser	cct Pro	2592
ggc Gly 865	gcc Ala	ctg Leu	gtc Val	gtc Val	ggg Gly 870	gtc Val	gtg Val	tgt Cys	gca Ala	gca Ala 875	ata Ile	ctg Leu	cgt Arg	cga Arg	cac His 880	2640
gtg Val	ggt Gly	ccg Pro	gga Gly	gag Glu 885	Gly	gct Ala	gtg Val	cag Gln	tgg Trp 890	Met	aac Asn	cgg Arg	ctg Leu	ata Ile 895	Ala	2688
ttc Phe	gcc Ala	tcg Ser	cgg Arg 900	ggt Gly	aat Asn	cat His	gtt Val	tcc Ser 905	Pro	acg Thr	cac His	tat Tyr	gtg Val 910	Pro	gag Glu	2736

Ser			gca Ala													2784
			ctg Leu													2832
aca Thr 945	ccg Pro	tgt Cys	tcc Ser	ggc	tcg Ser 950	tgg Trp	cta Leu	agg Arg	gat Asp	gtt Val 955	tgg Trp	gac Asp	tgg Trp	ata Ile	tgc Cys 960	2880
			act Thr													2928
cag Gln																2976
gtc Val	tgg Trp	cgg Arg 995	Gly	gac Asp	ggc Gly	atc Ile	atg Met 1000	Gln	acc Thr	acc Thr	tgc Cys	cca Pro 1009	Cys	gga Gly	gca Ala	3024
cag Gln	atc Ile 1010	Thr	gga Gly	cat His	gtc Val	aaa Lys 1015	Asn	ggt Gly	tcc Ser	atg Met	agg Arg 1020	Ile	gtc Val	ggg Gly	cct Pro	3072
aad																
Lys 1025	Thr	tgc Cys	agc Ser	aac Asn	acg Thr 1030	Trp	cat His	gga Gly	aca Thr	Phe	Pro	atc Ile	aac Asn	gca Ala	tac Tyr 1040	3120
Lys	Thr	Cys	Ser	Asn tgc	Thr 1030 aca Thr	Trp) ccc	His tct	Gly	Thr gcg	Phe 1035 cca Pro	Pro aac	Ile tat	Asn	Ala agg	Tyr 1040 gcg Ala	3120
Lys 1025 acc	Thr acg Thr	cgg Gly	Ser ccc Pro	tgc Cys 1045 gcc Ala	Thr 1030 aca Thr	Trp) ccc Pro	tct Ser	CCa Pro	gcg Ala 1050 gtg Val	Phe 1035 cca Pro	Pro aac Asn	Ile tat Tyr acg	Asn tct Ser	agg Arg 1055 gtg Val	Tyr 1040 gcg Ala ggg	
Lys 1025 acc Thr	Thr acg Thr tgg Trp ttc Phe	ggc Gly cgg Arg	ccc Pro gtg Val 1060 tac	tgc Cys 1045 gcc Ala gtg Val	Thr 1030 aca Thr gct Ala acg Thr	Trp) ccc Pro gag Glu ggc Gly	tct Ser gag Glu atg Met	Cly cca Pro tac Tyr 1065	gcg Ala 1050 gtg Val act Thr	Phe 1035 cca Pro gag Glu gac Asp	aac Asn gtc Val	tat Tyr acg Thr	tct Ser cgg Arg 1070	agg Arg 1055 gtg Val	Tyr 1040 gcg Ala ggg Gly	3168
Lys 1025 acc Thr ctg Leu gat Asp	Thr acg Thr tgg Trp ttc Phe	ggc Gly cgg Arg cac His 1075	ccc Pro gtg Val 1060 tac Tyr	tgc Cys 1045 gcc Ala) gtg Val	Thr 1030 aca Thr gct Ala acg Thr	Trp Ccc Pro gag Glu ggc Gly	tct Ser gag Glu atg Met 1080 ttc Phe	CCA Pro tac Tyr 1065 acc Thr	gcg Ala 1050 gtg Val act Thr	Phe 1035 cca Pro gag Glu gac Asp	aac Asn gtc Val aac Asn gtg	tat Tyr acg Thr gta Val 1085 gac Asp	tct Ser cgg Arg 1070 aag Lys	agg Arg 1055 gtg Val tgc Cys	gcg Ala ggg Gly cca Pro	3168 3216
Lys 1025 acc Thr ctg Leu gat Asp	Thr acg Thr tgg Trp ttc Phe cag Gln 1090 cac His	ggc Gly cgg Arg cac His 1075 gtt Val	ccc Pro gtg Val 1060 tac Tyr ccg Pro	tgc Cys 1045 gcc Ala ytg Val	Thr 1030 aca Thr gct Ala acg Thr	Trp CCC Pro gag Glu ggc Gly gaa Glu 1095 gcg Ala	tct Ser gag Glu atg Met 1080 ttc Phe	CCa Pro tac Tyr 1065 acc Thr	gcg Ala 1050 gtg Val act Thr acg	Phe 1035 cca Pro gag Glu gac Asp	aac Asn gtc Val aac Asn gtg Val 1100 cta Leu	tat Tyr acg Thr gta Val 1085 gac Asp	tct Ser cgg Arg 1070 aag Lys gga Gly	agg Arg 1055 gtg Val tgc Cys gtg Val	gcg Ala ggg Gly cca Pro cgg Arg	3168 3216 3264

tgc gag co Cys Glu P	cc gaa o ro Glu 1 1140	ccg gat Pro Asp	gta gc Val Al	a gtg a Val 1145	Leu '	act t Thr S	cc a Ser M	1et	ctc Leu 1150	Thr	gac Asp	3456
ccc tcc cc Pro Ser H	ac atc a is Ile '	aca gca Thr Ala	gaa ac Glu Th 11	r Ala	aag (Lys 2	cgt a Arg A	\rg I	tg Leu 1165	Ala	agg Arg	GJA āāā	3504
tct ccc cc Ser Pro P: 1170	cc tcc 1 ro Ser 1	Leu Ala	agc tc Ser Se 1175	t tca r Ser	gct (Ser G	ag t Eln I 180	ttg Leu	tct Ser	gcg Ala	cct Pro	3552
tcc ttg ad Ser Leu Ly 1185	ag gcg a ys Ala '	aca tgc Thr Cys 1190	Thr Th	c cac r His	His '	gtc t Val S 1195	ct o Ser I	ccg Pro	gac Asp	Ala	gac Asp 1200	3600
ctc atc galle G	lu Ala 2	aac ctc Asn Leu 1205	ctg tg Leu Tr	g cgg p Arg	cag Gln 1210	gag a Glu M	atg (Met (ggc Gly	GJÀ aaa	aac Asn 1215	11e	3648
acc cgc g Thr Arg V	tg gag al Glu 1220	tcg gag Ser Glu	aac aa Asn Ly	g gtg s Val 1225	Val	gtc c Val I	ctg (Leu <i>l</i>	gac Asp	tct Ser 1230	Phe	gac Asp	3696
ccg ctt c Pro Leu A 1	ga gcg rg Ala (235	gag gag Glu Glu	Asp Gl	g agg u Arg 40	gaa Glu	gta t Val S	Ser V	gtt Val 1245	Pro	gcg Ala	gag Glu	3744
atc ctg c Ile Leu A 1250	gg aaa rg Lys	tcc aag Ser Lys	aag tt Lys Ph 1255	c ccc e Pro	gca Ala	Ala N	atg d Met 1 1260	Pro	atc Ile	tgg Trp	gcg Ala	3792
Ile Leu A	rg Lys	Ser Lys	Lys Ph 1255 cca ct Pro Le	e Pro	Ala gag Glu	Ala N 1 tcc t	Met 1 1260 tgg a	Pro aag	Ile gac	Trp	gac	3792
lle Leu A 1250 cgc ccg g Arg Pro A	at tac sp Tyr cet ecg	Ser Lys aac cct Asn Pro 1270	Lys Ph 1255 cca ct Pro Le	g tta u Leu	gag Glu	tcc t Ser 7 1275 ttg (Met 1 1260 tgg a Trp 1	Pro aag Lys	gac Asp	CCG Pro	gac Asp 1280 gcc Ala	
cgc ccg g Arg Pro A 1265	at tac sp Tyr sct ccg Pro Pro	Ser Lys aac cct Asn Pro 1270 gtg gtg Val Val 1285	Lys Ph 1255 cca ct Pro Le cac gg His Gl cgg ag	g tta u Leu g tgc y Cys a aag	gag Glu ccg Pro 1290 agg Arg	tcc t Ser 1 1275 ttg (Leu 1	Met 1 1260 tgg (Trp) cca (Pro)	aag Lys cct Pro	gac Asp atc Ile	ccg Pro aag Lys 129	gac Asp 1280 gcc Ala 5	3840
Ile Leu A 1250 cgc ccg g Arg Pro A 1265 tac gtc c Tyr Val P cct cca a Pro Pro I tcc tcc g Ser Ser V	at tac sp Tyr ct ccg Pro Pro	Ser Lys aac cct Asn Pro 1270 gtg gtg Val Val 1285 cct cca Pro Pro	cca ct Pro Le cac gg His Gl cgg ag Arg Ar	g tta u Leu g tgc y Cys a aag g Lys 130	gag Glu ccg Pro 1290 agg Arg	tcc t Ser 7 1275 ttg (Leu)	Met 11260 tgg a Trp 1 cca a Pro 1 gtt a Val a act	aag Lys cct Pro gtc Val	gac Asp atc Ile cta Leu 1310	ccg Pro aag Lys 1299 aca Thr	gac Asp 1280 gcc Ala 5 gag Glu	3840
Ile Leu A 1250 cgc ccg g Arg Pro A 1265 tac gtc c Tyr Val P cct cca a Pro Pro I tcc tcc g Ser Ser V	at tac sp Tyr ct ccg ro Pro ata cca le Pro 1300 stg tct val Ser 315	aac cct Asn Pro 1270 gtg gtg Val Val 1285 cct cca Pro Pro tct gcc Ser Ala	Lys Ph 1255 cca ct Pro Le cac gg His Gl cgg ag Arg Ar tta gc Leu Al 13	g tta u Leu g tgc y Cys a aag g Lys 1300 g gag a Glu 220	gag Glu ccg Pro 1290 agg Arg ctc Leu	tcc t Ser 7 1275 ttg c Leu 1 acg c Thr v	Met 11260 tgg 6 Trp 1 cca 6 Pro 1 gtt 7 Val 7 act Thr	Pro aag Lys cct Pro gtc Val aag Lys 1325	gac Asp atc Ile cta Leu 1310 acc Thr	ccg Pro aag Lys 1299 aca Thr ttc Phe	gac Asp 1280 gcc Ala 5 gag Glu	3840

Asp 134		Ala	Ser	Asp	Asp 135		Asp	Lys	Gly	Ser 135		Val	Glu	Ser	туг 1360	
		_			Leu					Gly				ctc Leu 137	Ser	4128
				Ser					Glu					gtc Val O		4176
			Met					Thr					Thr	cca Pro		4224
		Glu					Pro					Ser		tct Ser		4272
	Arg					Val					Ser			gca Ala		4320
ctg Leu	cgg Arg	cag Gln	aag Lys	aag Lys 1445	Val	acc Thr	ttt Phe	gac Asp	aga Arg 1450	Leu	caa Gln	gtc Val	ctg Leu	gac Asp 1455	Asp	4368
				Val					Lys					aca Thr		4416
			Leu					Glu					Thr	ccc Pro		4464
		Ala					Gly					Asp		cgg Arg		4512
	Ser					Asn					Val			gac Asp		4560
					Thr					Thr				aaa Lys 1535	Asn	4608
	_		-	Val			Glu		Gly		_	_		gcc Ala	-	4656
			Phe					Val					Lys	atg Met		4704

ctc Leu	tat Tyr 1570	Asp	gtg Val	gtc Val	tcc Ser	acc Thr 1575	Leu	cct Pro	cag Gln	gtc Val	gtg Val 1580	met	ggc Gly	tcc Ser	tca Ser	4752
tac Tyr 1585	Gly	ttc Phe	cag Gln	tac Tyr	tct Ser 1590	Pro	Gly ggg	cag Gln	cga Arg	gtc Val 1595	GIu	ttc Phe	ctg Leu	gtg Val	aat Asn 1600	4800
acc Thr	tgg Trp	aaa Lys	tca Ser	aag Lys 1605	Lys	aac Asn	ccc Pro	atg Met	ggc Gly 1610	Phe	tca Ser	tat Tyr	gac Asp	act Thr 1615	Arg	4848
tgt Cys	ttc Phe	gac Asp	tca Ser 1620	Thr	gtc Val	acc Thr	gag Glu	aac Asn 1625	Asp	atc Ile	cgt Arg	gtt Val	gag Glu 1630	GIU	tca Ser	4896
att Ile	tac Tyr	caa Gln 163	Cys	tgt Cys	gac Asp	ttg Leu	gcc Ala 1640	Pro	gaa Glu	gcc Ala	aga Arg	cag Gln 1645	Ala	ata Ile	aaa Lys	4944
tcg Ser	ctc Leu 1650	Thr	gag Glu	cgg Arg	ctt Leu	tat Tyr 165	Ile	Gly	ggt Gly	cct Pro	ctg Leu 166	Thr	aat Asn	tca Ser	aaa Lys	4992
ggg Gly 1665	Gln	aac Asn	tgc Cys	ggt Gly	tat Tyr 167	Arg	cgg Arg	tgc Cys	cgc Arg	gcg Ala 167	Ser	ggc Gly	gtg Val	ctg Leu	acg Thr 1680	5040
act Thr	agc Ser	tgc Cys	ggt Gly	aac Asn 168	Thr	ctc Leu	aca Thr	tgt Cys	tac Tyr 169	Leu	aag Lys	gcc Ala	tct Ser	gca Ala 169	Ala	5088
tgt Cys	cga Arg	gct Ala	gcg Ala 170	Lys	ctc Leu	cag Gln	gac Asp	tgc Cys 170	Thr	atg Met	ctc Leu	gtg Val	aac Asn 171	GTA	gac Asp	5136
gac Asp	ctt Leu	gtc Val 171	Val	atc Ile	tgt Cys	gaa Glu	agc Ser 172	Ala	gga Gly	acc Thr	caa Gln	gag Glu 172	Asp	gcg Ala	gcg Ala	5184
agc Ser	cta Leu 173	Arg	gtc Val	Phe	Thr	gag Glu 173	Ala	Met	Thr	Arg	. JAI	Ser	gcc Ala	Pro	ccc Pro	5232
ggg Gly 174	Asp	ccg Pro	ccc Pro	caa Gln	cca Pro 175	Glu	tac Tyr	gac Asp	ttg Leu	gag Glu 175	Lev	ata Ile	aca Thr	tca Ser	tgt Cys 1760	5280
tcc Ser	tcc Ser	aat Asn	gtg Val	tcg Ser 176	Val	gcc Ala	cac	gat Asp	gca Ala 177	Ser	ggc Gly	: aaa / Lys	agg Arg	gtg Val 177	tac Tyr 5	5328
tac Tyr	cto	acc Thr	cgt Arg 178	Asp	ccc Pro	e acc	acc Thr	2 CCC Pro 178	Let	gca Ala	a cgg	g gct g Ala	gcg Ala 179	TIL	gaa Glu	5376

aca gct aga cac act cca gtt aac tcc tgg cta ggc aac att atc atg 5424 Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met 1800 1795 tat gcg ccc act ttg tgg gca agg atg att ctg atg act cac ttc ttc 5472 Tyr Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe 1810 1815 tcc atc ctt cta gca cag gag caa ctt gaa aaa gcc ctg gac tgc cag 5520 Ser Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln 1835 1830 1825 atc tac ggg gcc tgt tac tcc att gag cca ctt gac cta cct cag atc 5568 Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile 1850 att gaa cga ctc cat ggc ctt agc gca ttt tca ctc cat agt tac tct 5616 Ile Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser 1865 1860 cca ggt gag atc aat agg gtg gct tca tgc ctc agg aaa ctt ggg gta 5664 Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val 1880 1885 cca ccc ttg cga gtc tgg aga cat cgg gcc agg agc gtc cgc gct agg 5712 Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg 1900 1895 1890 cta ctg tcc cag ggg ggg agg gcc gcc act tgt ggc aag tac ctc ttc 5760 Leu Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe 1905 1910 1915 5808 aac tgg gca gtg aag acc aaa ctc aaa ctc act cca atc ccg gct gcg Asn Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala 1930 1925 tcc cag ctg gac ttg tcc ggc tgg ttc gtt gct ggt tac agc ggg gga 5856 Ser Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly 1940 1945 5904 gac ata tat cac agc ctg tct cgt gcc cga ccc cgc tgg ttc atg ctg Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu 1955 1960 tgc cta ctc cta ctt tct gta ggg gta ggc atc tac ctg ctc ccc aac 5952 Cys Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn 1975 1980 5955 cga Arg 1985 <210> 6 <211> 1984

```
<212> PRT
<213> Artificial Sequence
<220>
<223> NS sequence
<400> 6
Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys
                             10
Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu
                           25
Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val
                       40
Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu
                 55
Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln
                              75
                70
Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro
                            90
             85
Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp
                                110
              105
          100
Val Ile Pro Val Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser
                                125
           120
      115
Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu
                            140
       135
Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr
                       155
      150
Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu
                    170
                                               175
            165
Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala
                                         190
         180 185
Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser
                                        205
   195 200
Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys
                                     220
   210 215
Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala
                                 235
               230
Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val
                              250
Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly Lys
                                   270
                          265
Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile
                                       285
                       280
       275
Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly
                              300
                    295
Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu
                                 315
                 310
Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile
                                               335
                             330
Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys
                                         350
                  345
Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe Cys
                                     365
       355 360
His Ser Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu
                                    380
                    375
```

```
Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile
                                    395
Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr
              405
                                410
Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val
                   425
           420
Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr
                440
Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Arg
                                     460
                     455
Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu
                 470
                                    475
Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp
              485
                                490
Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg
                             505
                                                510
Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His
                                            525
                          520
Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala
                      535
                                        540
His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu
                                     555
                  550
Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro
                                570
              565
Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu
                             585
           580
His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu
                         600
                                            605
Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser
                     615
                                        620
Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val
                                    635
                  630
Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val Ile
              645
                                650
Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp Arg
                            665
                                               670
Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser His
                         680
                                           685
Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys Gln
                                       700
                     695
Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala Ala
                                    715
                 710
Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp Ala
                                730
              725
Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu
          740
                             745
Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr
                         760
Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe Asn
                      775
                                        780
Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala Ala
                 790
                                    795
Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser Ile
                                810
              805
Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly
```

			820					825					830		
t/a l	Δla	Glv	Ala	Leu	Val	Ala	Phe	Lys	Val	Met	Ser	Gly	Glu	Met	Pro
		235					840					845			
Ser	Thr	Glu	Asp	Leu	Val	Asn	Leu	Leu	Pro	Ala	Ile	Leu	Ser	Pro	GIA
	050					855					860				
	Leu	Val	Val	Gly	Val 870	Vai	Cys	Ala	AId	875	Leu	Arg	nr 9		880
865	Dro	Clv	Clu	Gly	Ala	Val	Gln	Trp	Met		Arg	Leu	Ile	Ala	Phe
				225					890					923	
Ala	Ser	Arg	Gly	Asn	His	Val	Ser	Pro	Thr	His	Tyr	Val	Pro	Glu	Ser
			900					905					210		
Asp	Ala		Ala	Arg	Val	Thr	Gln	Ile	Leu	Ser	Ser	ьец 925	THE	TIE	IIII
~1	T	915	T	7~~	Leu	Wie	920 Gln	Tro	Tle	Asn	Glu		Cys	Ser	Thr
	930					935					940				
Pro	Cys	Ser	Gly	Ser	Trp	Leu	Arg	Asp	Val	Trp	Asp	\mathtt{Trp}	Ile	Cys	Thr
DAE					950					955					900
Val	Leu	Thr	Asp		Lys	Thr	Trp	Leu	970	Ser	ьys	Leu	reu	975	GIII
•	D	~ 1	1707	965	Phe	Phe	Ser	Cvs		Ara	Glv	Tyr	Lys		Val
			980					985					990		
Trp	Arg	Gly	Asp	Gly	Ile	Met	Gln	Thr	Thr	Cys	Pro	Cys	Gly	Ala	Gln
		995					100	יט				TOO	9		
Ile			His	Val	Lys	Asn	Gly	Ser	Met	Arg	11e 102	o Vai	GIY	Pro	rys
	1010	0	•	m>	Trp	1015) (1)	Thr	Phe	Pro	Tle	Asn	Ala	Tvr	Thr
102	E				1030)				702	-				10-10
Thr	Glv	Pro	Cvs	Thr	Pro	Ser	Pro	Ala	Pro	Asn	Tyr	Ser	Arg	Ala	Leu
				104	5				105	U				100	•
Trp	Arg	Val			Glu	Glu	Tyr	Val	Glu	Val	Thr	Arg	107	U GTĀ	Asp
		_	106	0 ———	Gly	Mot	Th~	106) Den	Δen	Val	Lvs			Cys
Phe	His	Tyr 107		Thr	GIY	Met	108	0	nsp	71011	,	108	5		•
Gln	Va1	Pro	Ala	Pro	Glu	Phe	Phe	Thr	Glu	Val	Asp	Gly	Val	Arg	Leu
	100	Λ				109	5				110	U			
His	Arg	Tyr	Ala	Pro	Ala	Cys	Arg	Pro	Leu	Leu	Arg	Glu	Glu	Val	1120
110	5			_	111	0			. 17-1	111		Gln	I.e.	Pro	
Phe	Gln	Val	Gly	Leu 112	Asn	GIN	TYL	neo	113	. GIY	361	01	Dea	113	5
C311	Pro	Glu	Pro	ASD	Val	Ala	Val	Leu			Met	Leu	Thr	Asp	Pro
			114	.n				114	5				TTO	v	
Ser	His	Ile	Thr	Ala	Glu	Thr	Ala	Lys	Arc	Arg	Leu	Ala	Arg	Gly	Ser
		115	5		_	_	116	0			Tax	116		Pro	Ser
Pro	Pro	Ser	Leu	Ala	Ser	Ser 117	Ser 5	Ala	Ser	. G11	118	0	ALG		Ser
7 011	117	ังไล	ጥኮተ	· Cvs	Thr	Thr	His	His	. Val	Ser			Ala	Asp	Leu
110	5				119	0				119	15				1200
Ile	Glu	Ala	Asn	Lev	Leu	Trp	Arc	, Glr	ı Glı	Met	: Gly	r Gly	Asn	Ile	Thr
				120	15				123	LO				121	
Arg	val	Glu			Asn	Lys	Val	. Va. 122	L Val	r rer	ı ASĮ	, ser	123	: vəř	Pro
	. x	. או	122	(U 1 (C) 1) Der	, Glu	Arc	124 Gli	ر. Va`	l Ser	. Val	Pro			Ile
		127	3.5				124	10				124	10		
Let	Arc	Lys	s Sei	Lys	. Lys	Phe	Pro	Ala	a Ala	a Met	Pro	Ile	Trp	Ala	Arg
	125					125	5				126	0			

Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro Asp Tyr 1270 1275 Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala Pro 1285 1290 Pro Ile Pro Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu Ser 1300 1305 1310 Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly Ser 1315 1320 1325 Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro Asp 1330 1335 1340 Gln Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr Ser 1345 1350 1355 1360 Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp 1365 1370 1375 Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val Cys 1380 1385 1390 Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala 1395 1400 1405 Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu 1410 1415 1420 Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly Leu 1430 1435 Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp His 1445 1450 1455 Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys 1460 1465 1470 Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro His 1475 1480 1485 Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu 1490 1495 1500 Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu Leu 1510 1515 1520 1505 Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu 1525 1530 1535 Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu 1545 1540 Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu 1555 1560 1565 Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser Tyr 1570 1575 1580 Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn Thr 1590 1595 1600 Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys 1605 1610 1615 Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser Ile 1620 1625 1630 Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys Ser 1635 1640 1645 Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly 1650 1655 1660 Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr 1665 1670 1675 1680 Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala Cys 1685 1690

```
Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Gly Asp Asp
                                 1710
                        1705
         1700
Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala Ser
      1715 1720 1725
Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly
  1730 1735 1740
Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser
                                                1760
1745 1750 1755
Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr Tyr
            1765 1770 1775
Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr
                       1785 · 1790
         1780
Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Tyr
      1795 1800 1805
Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser
                1815 1820
Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln Ile
       1830 1835
Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile Ile
            1845 1850 1855
Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro
       1860 1865 1870
Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val Pro
     1875 1880
                                      1885
Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu
          1895
                                 1900
Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe Asn
        1910
                                1915
Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala Ser
            1925
                             1930
Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly Asp
                         1945
         1940
Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu Cys
     1955 1960 1965
Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
                                    1980
                    1975
   1970
<210> 7
<211> 4909
<212> DNA
<213> Artificial Sequence
<220>
<223> pVlJ nucleic acid
tegegegttt eggtgatgae ggtgaaaace tetgacacat geageteeeg gagaeggtea
                                                          60
cagettgtet gtaageggat geegggagea gacaageeeg teagggegeg teagegggtg
                                                          120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
                                                          180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg
                                                          240
ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg
                                                          300
tecaacatta cegecatgtt gacattgatt attgactagt tattaatagt aatcaattae
                                                          360
                                                          420
ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg
cccgcctggc tgaccgcca acgacccccg cccattgacg tcaataatga cgtatgttcc
                                                          480
catagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac
                                                          540
```

tgcccacttg	gcagtacatc	aagtgtatca	tatgccaagt	acgcccccta	ttgacgtcaa	600
tgacggtaaa	tggcccgcct	ggcattatgc	ccagtacatg	accttatggg	actttcctac	660
				gtgatgcggt		720
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	accccattga	780
				tttccaaaat		840
				tgggaggtct		900
agctcgttta	gtgaaccgtc	agategeetg	gagacgccat	ccacgctgtt	ttgacctcca	960
tagaagacac	cgggaccgat	ccagcctccg	cggccgggaa	cggtgcattg	gaacgcggat	1020
tccccgtgcc	aagagtgacg	taagtaccgc	ctatagactc	tataggcaca	cccctttggc	1080
tcttatgcat	gctatactgt	ttttggcttg	gggcctatac	acccccgctt	ccttatgcta	1140
taggtgatgg	tatagcttag	cctataggtg	tgggttattg	accattattg	accactcccc	1200
				ctctttgcca		1260
				acggactctg		1320
				acaacgccgt		1380
				tctcgggtac		1440
				agccctggtc		1500
				tggaggccag		1560
				tggcggtagg		1620
				atggaagact		1680
gcagaagaag	atgcaggcag	ctgagttgtt	gtattctgat	aagagtcaga	ggtaactccc	1740
gttgcggtgc	tgttaacggt	ggagggcagt	gtagtctgag	cagtactcgt	tgctgccgcg	1800
cgcgccacca	gacataatag	ctgacagact	aacagactgt	tcctttccat	gggtcttttc	1860
tgcagtcacc	gtccttagat	ctaggtacca	gatatcagaa	ttcagtcgac	agcggccgcg	1920
				ccctccccg		1980
				aatgaggaaa		2040
				gggcaggaca		2100
				ggctctatgg		2160
caggtgctga	agaattgacc	cggttcctcc	tgggccagaa	agaagcaggc	acatcccctt	2220
ctctgtgaca	caccctgtcc	acgcccctgg	ttcttagttc	cagccccact	cataggacac	2280
tcatagctca	ggagggctcc	gccttcaatc	ccacccgcta	aagtacttgg	agcggtctct	2340
				gagtgggaag		2400
				aacatgtgag		2460
				tcgctgcgct		2520
				cggttatcca		2580
				aaggccagga		2640
ggccgcgttg	ctggcgtttt	tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	2700
				agataccagg		2760
tggaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgccg	cttaccggat	acctgtccgc	2820
ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	2880
ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	cccccgttc	agcccgaccg	2940
				gtaagacacg		3000
				tatgtaggcg		3060
gttcttgaag.	tggtggccta	actacggcta	cactagaaga	acagtatttg	gtatctgcgc	3120
tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	3180
caccgctggt	agcggtggtt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	3240
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaactc	3300
acgttaaggg	attttggtca	tgagattatc	aaaaaggatc	ttcacctaga	tccttttaaa	3360
ttaaaaatga	agttttaaat	caatctaaag	tatatatgag	taaacttggt	ctgacagtta	3420
ccaatgctta	atcagtgagg	cacctatctc	agcgatctgt	ctatttcgtt	catccatagt	3480
tgcctgactc	ggggggggg	ggcgctgagg	tctgcctcgt	gaagaaggtg	ttgctgactc	3540
ataccaggcc	tgaatcgccc	catcatccag	ccagaaagtg	agggagccac	ggttgatgag	3600
agctttgttg	taggtggacc	agttggtgat	tttgaacttt	tgctttgcca	cggaacggtc	3660
tgcgttgtcg	ggaagatgcg	tgatctgatc	cttcaactca	gcaaaagttc	gatttattca	3720
acaaagccgc	cgtcccgtca	agtcagcgta	atgctctgcc	agtgttacaa	ccaattaacc	3780
aattctgatt	agaaaaactc	atcgagcatc	aaatgaaact	gcaatttatt	catatcagga	3840

```
ttatcaatac catatttttg aaaaagccgt ttctgtaatg aaggagaaaa ctcaccgagg
                                                                     3900
cagttecata ggatggcaag atcetggtat eggtetgega tteegaeteg tecaacatea
                                                                     3960
                                                                     4020
atacaaccta ttaatttccc ctcgtcaaaa ataaggttat caagtgagaa atcaccatga
gtgacgactg aatccggtga gaatggcaaa agcttatgca tttctttcca gacttgttca
                                                                     4080
                                                                     4140
acaggecage cattaegete gteateaaaa teactegeat caaccaaace gttatteatt
cgtgattgcg cctgagcgag acgaaatacg cgatcgctgt taaaaggaca attacaaaca
                                                                     4200
ggaatcgaat gcaaccggcg caggaacact gccagcgcat caacaatatt ttcacctgaa
                                                                     4260
tcaggatatt cttctaatac ctggaatgct gttttcccgg ggatcgcagt ggtgagtaac
                                                                     4320
catgcatcat caggagtacg gataaaatgc ttgatggtcg gaagaggcat aaattccgtc
                                                                     4380
agccagttta gtctgaccat ctcatctgta acatcattgg caacgctacc tttgccatgt
                                                                     4440
ttcagaaaca actctggcgc atcgggcttc ccatacaatc gatagattgt cgcacctgat
                                                                     4500
tgcccgacat tatcgcgagc ccatttatac ccatataaat cagcatccat gttggaattt
                                                                     4560
aatcgcggcc tcgagcaaga cgtttcccgt tgaatatggc tcataacacc ccttgtatta
                                                                     4620
ctgtttatgt aagcagacag ttttattgtt catgatgata tatttttatc ttgtgcaatg
                                                                     4680
taacatcaga gattttgaga cacaacgtgg ctttcccccc cccccatta ttgaagcatt
                                                                     4740
tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa aaataaacaa
                                                                     4800
ataggggttc cgcgcacatt tccccgaaaa gtgccacctg acgtctaaga aaccattatt
                                                                     4860
                                                                     4909
atcatgacat taacctataa aaataggcgt atcacgaggc cctttcgtc
<210> 8
<211> 35935
<212> DNA
<213> Adenovirus serotype 6
<400> 8
catcatcaat aatatacctt attttggatt gaagccaata tgataatgag ggggtggagt
                                                                       60
ttgtgacgtg gcgcggggcg tgggaacggg gcgggtgacg tagtagtgtg gcggaagtgt
                                                                      120
                                                                      180
gatgttgcaa gtgtggcgga acacatgtaa gcgacggatg tggcaaaagt gacgtttttg
gtgtgcgccg gtgtacacag gaagtgacaa ttttcgcgcg gttttaggcg gatgttgtag
                                                                      240
taaatttggg cgtaaccgag taagatttgg ccattttcgc gggaaaactg aataagagga
                                                                      300
agtgaaatct gaataatttt gtgttactca tagcgcgtaa tatttgtcta gggccgcggg
                                                                      360
                                                                      420
gactttgacc gtttacgtgg agactcgccc aggtgttttt ctcaggtgtt ttccgcgttc
                                                                       480
cgggtcaaag ttggcgtttt attattatag tcagctgacg tgtagtgtat ttatacccgg
                                                                       540
tgagttcctc aagaggccac tcttgagtgc cagcgagtag agttttctcc tccgagccgc
tccgacaccg ggactgaaaa tgagacatat tatctgccac ggaggtgtta ttaccgaaga
                                                                       600
aatggccgcc agtcttttgg accagctgat cgaagaggta ctggctgata atcttccacc
                                                                       660
tcctagccat tttgaaccac ctacccttca cgaactgtat gatttagacg tgacggcccc
                                                                       720
cgaagatccc aacgaggagg cggtttcgca gatttttccc gactctgtaa tgttggcggt
                                                                       780
gcaggaaggg attgacttac tcacttttcc gccggcgccc ggttctccgg agccgcctca
                                                                       840
                                                                       900
cetttecegg cageeegage ageeggagea gagageettg ggteeggttt etatgeeaaa
                                                                       960
ccttgtaccg gaggtgatcg atcttacctg ccacgaggct ggctttccac ccagtgacga
cgaggatgaa gagggtgagg agtttgtgtt agattatgtg gagcaccccg ggcacggttg
                                                                      1020
caggictigt cattatcacc ggaggaatac gggggaccca gatattatgt gttcgctttg
                                                                      1080
ctatatgagg acctgtggca tgtttgtcta cagtaagtga aaattatggg cagtgggtga
                                                                      1140
tagagtggtg ggtttggtgt ggtaattttt tttttaattt ttacagtttt gtggtttaaa
                                                                      1200
gaattttgta ttgtgatttt tttaaaaggt cctgtgtctg aacctgagcc tgagcccgag
                                                                      1260
ccagaaccgg agcctgcaag acctacccgc cgtcctaaaa tggcgcctgc tatcctgaga
                                                                      1320
cgcccgacat cacctgtgtc tagagaatgc aatagtagta cggatagctg tgactccggt
                                                                      1380
cettetaaca caceteetga gatacacecg gtggteeege tgtgceecat taaaccagtt
                                                                      1440
gccgtgagag ttggtgggcg tcgccaggct gtggaatgta tcgaggactt gcttaacgag
                                                                      1500
                                                                      1560
cctgggcaac ctttggactt gagctgtaaa cgccccaggc cataaggtgt aaacctgtga
                                                                      1620
ttgcgtgtgt ggttaacgcc tttgtttgct gaatgagttg atgtaagttt aataaagggt
gagataatgt ttaacttgca tggcgtgtta aatggggcgg ggcttaaagg gtatataatg
                                                                      1680
cgccgtgggc taatcttggt tacatctgac ctcatggagg cttgggagtg tttggaagat
                                                                      1740
                                                                      1800
ttttctgctg tgcgtaactt gctggaacag agctctaaca gtacctcttg gttttggagg
```

tttctgtggg	gctcatccca	ggcaaagtta	gtctgcagaa	ttaaggagga	ttacaagtgg	1860
gaatttgaag	agcttttgaa	atcctgtggt	gagctgtttg	attctttgaa	tctgggtcac	1920
caggcgcttt	tccaagagaa	ggtcatcaag	actttggatt	tttccacacc	ggggcgcgct	1980
acaactacta	ttgcttttt	gagttttata	aaggataaat	ggagcgaaga	aacccatctg	2040
agcgggggt	acctgctgga	ttttctggcc	atgcatctgt	ggagagcggt	tgtgagacac	2100
aagaatcgcc	tactactatt	gtcttccgtc	cgcccggcga	taataccgac	ggaggagcag	2160
cagcagcagc	aggaggaagc	caggcggcgg	cggcaggagc	agagcccatg	gaacccgaga	2220
accaacctaa	accetegga	atgaatgttg	tacaggtggc	tgaactgtat	ccagaactga	2280
gacgcatttt	gacaattaca	gaggatgggc	aggggctaaa	gggggtaaag	agggagcggg	2340
gggcttgtga	ggctacagag	gaggctagga	atctagcttt	tagcttaatg	accagacacc	2400
gtcctgagtg	tattactttt	caacagatca	aggataattg	cgctaatgag	cttgatctgc	2460
tagcacagaa	gtattccata	gagcagctga	ccacttactg	gctgcagcca	ggggatgatt	2520
ttgaggaggc	tattagggta	tatgcaaagg	tggcacttag	gccagattgc	aagtacaaga	2580
tcagcaaact	totaaatatc	aggaattgtt	gctacatttc	tgggaacggg	gccgaggtgg	2640
agatagatac	ggaggatagg	gtggccttta	gatgtagcat	gataaatatg	tggccggggg	2700
tacttagcat	ggacggggtg	gttattatga	atgtaaggtt	tactggcccc	aattttagcg	2760
gtacggtttt	cctggccaat	accaacctta	tcctacacgg	tgtaagcttc	tatgggttta	2820
acaataccto	tatagaagcc	tggaccgatg	taagggttcg	gggctgtgcc	ttttactgct	2880
gctggaaggg	agtagtatat	cgccccaaaa	gcagggcttc	aattaagaaa	tgcctctttg	2940
aaaggtgtac	cttgggtatc	ctgtctgagg	gtaactccag	ggtgcgccac	aatgtggcct	3000
ccgactgtgg	ttgcttcatg	ctagtgaaaa	gcgtggctgt	gattaagcat	aacatggtat	3060
gtggcaactg	cgaggacagg	gcctctcaga	tgctgacctg	ctcggacggc	aactgtcacc	3120
toctgaagac	cattcacgta	gccagccact	ctcgcaaggc	ctggccagtg	tttgagcata	3180
acatactgac	ccactattcc	ttgcatttgg	gtaacaggag	gggggtgttc	ctaccttacc	3240
aatgcaattt	gagtcacact	aagatattgc	ttgagcccga	gagcatgtcc	aaggtgaacc	3300
tgaacggggt	gtttgacatg	accatgaaga	tctggaaggt	gctgaggtac	gatgagaccc	3360
gcaccaggtg	cagaccctgc	gagtgtggcg	gtaaacatat	taggaaccag	cctgtgatgc	3420
tagatataac	cgaggagctg	aggcccgatc	acttggtgct	ggcctgcacc	cgcgctgagt	3480
ttggctctag	cgatgaagat	acagattgag	gtactgaaat	gtgtgggcgt	ggcttaaggg	3540
toggaaagaa	tatataaggt	gggggtctta	tgtagttttg	tatctgtttt	gcagcagccg	3600
ccaccaccat	gagcaccaac	tcgtttgatg	gaagcattgt	gagctcatat	ttgacaacgc	3660
gcatgcccc	atgggccggg	gtgcgtcaga	atgtgatggg	ctccagcatt	gatggtcgcc	3720
ccatcctacc	cgcaaactct	actaccttga	cctacgagac	cgtgtctgga	acgccgttgg	3780
agactgcagc	ctccaccacc	gcttcagccg	ctgcagccac	cgcccgcggg	attgtgactg	3840
actttgcttt	cctgagcccg	cttgcaagca	gtgcagcttc	ccgttcatcc	gcccgcgatg	3900
acaagttgac	gactettttg	gcacaattgg	attetttgac	ccgggaactt	aatgtcgttt	3960
ctcagcagct	attagatcta	cgccagcagg	tttctgccct	gaaggcttcc	tcccctccca	4020
atgcggttta	aaacataaat	aaaaaaccag	actctgtttg	gatttggatc	aagcaagtgt	4080
cttactatct	ttatttaggg	gttttgcgcg	cgcggtaggc	ccgggaccag	cggtctcggt	4140
cattaagaat	cctqtqtatt	ttttccagga	cgtggtaaag	gtgactctgg	atgttcagat	4200
acatgggcat	aagcccqtct	ctggggtgga	ggtagcacca	ctgcagagct	tcatgctgcg	4260
agatagtatt	gtagatgatc	cagtcgtagc	aggagcgctg	ggcgtggtgc	ctaaaaatgt	4320
ctttcagtag	caagetgatt	gccaggggca	ggcccttggt	gtaagtgttt	acaaagcggt	4380
taagctggga	tgggtgcata	cgtggggata	tgagatgcat	cttggactgt	atttttaggt	4440
tagctatatt	cccaqccata	tccctccggg	gattcatgtt	gtgcagaacc	accagcacag	4500
tatatccagt	gcacttggga	aatttgtcat	gtagcttaga	aggaaatgcg	tggaagaact	4560
tagagacacc	cttgtgacct	ccaagatttt	ccatgcattc	gtccataatg	atggcaatgg	4620
acccacaaac	aacaacctaa	gcgaagatat	ttctgggatc	actaacgtca	tagttgtgtt	4680
		gccatttta				4740
gtataatggt	tccatccggc	ccaggggcgt	agttaccctc	acagatttgc	atttcccacg	4800
ctttgagttc	agatgggggg	atcatgtcta	cctgcggggc	gatgaagaaa	acggtttccg	4860
gggtaggga	gatcagctgg	gaagaaagca	ggttcctgag	cagctgcgac	ttaccgcagc	4920
caataaaccc	gtaaatcaca	cctattaccg	ggtgcaactg	gtagttaaga	gagctgcagc	4980
tgccgtcatc	cctgagcagg	ggggccactt	cgttaagcat	gtccctgact	cgcatgtttt	5040
ccctgaccaa	atccgccaga	aggcgctcgc	cgcccagcga	tagcagttct	tgcaaggaag	5100
-	- •	_		_		

caaaqttttt	caacggtttg	agaccgtccg	ccgtaggcat	gcttttgagc	gtttgaccaa	5160
gcagttccag	gcggtcccac	agctcggtca	cctgctctac	ggcatctcga	tccagcatat	5220
ctcctcattt	cacagattag	ggcggctttc	gctgtacggc	agtagtcggt	gctcgtccag	5280
acgggccagg	gtcatgtctt	tccacgggcg	cagggtcctc	gtcagcgtag	tctgggtcac	5340
ggtgaagggg	tacactccgg	gctgcgcgct	ggccagggtg	cgcttgaggc	tggtcctgct	5400
ggtgctgaag	cactaccagt	cttcgccctg	cgcgtcggcc	aggtagcatt	tgaccatggt	5460
greatagtee	agcccctccg	cggcgtggcc	cttggcgcgc	agcttgccct	tggaggaggc	5520
accacacaaa	gggcagtgca	gacttttgag	ggcgtagagc	ttgggcgcga	gaaataccga	5580
ttccggggag	taggcatccg	cgccgcaggc	cccgcagacg	gtctcgcatt	ccacgagcca	5640
ggtgagctct	aaccattcgg	ggtcaaaaac	caggtttccc	ccatgctttt	tgatgcgttt	5700
cttacctctq	gtttccatga	gccggtgtcc	acgctcggtg	acgaaaaggc	tgtccgtgtc	5760
cccqtataca	gacttgagag	gcctgtcctc	gagcggtgtt	ccgcggtcct	cctcgtatag	5820
aaactcggac	cactctgaga	caaaggctcg	cgtccaggcc	agcacgaagg	aggctaagtg	5880
ggagggtag	caatcattgt	ccactagggg	gtccactcgc	tccagggtgt	gaagacacat	5940
atcaccctct	tcggcatcaa	ggaaggtgat	tggtttgtag	gtgtaggcca	cgtgaccggg	6000
tattectaaa	ggggggctat	aaaagggggt	gggggcgcgt	tcgtcctcac	tetetteege	6060
atcoctotct	gcgagggcca	gctgttgggg	tgagtactcc	ctctgaaaag	cgggcatgac	6120
ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	ttgatattca	cctggcccgc	6180
gatgatacct	ttgagggtgg	ccgcatccat	ctggtcagaa	aagacaatct	ttttgttgtc	6240
aagcttggtg	gcaaacgacc	cgtagagggc	gttggacagc	aacttggcga	tggagcgcag	6300
aatttaattt	ttgtcgcgat	cggcgcgctc	cttggccgcg	atgtttagct	gcacgtattc	6360
gcgcgcaacg	caccqccatt	cgggaaagac	ggtggtgcgc	tcgtcgggca	ccaggtgcac	6420
gcgccaaccg	caattataca	gggtgacaag	gtcaacgctg	gtggctacct	ctccgcgtag	6480
acactcatta	gtccagcaga	ggcggccgcc	cttgcgcgag	cagaatggcg	gtagggggtc	6540
tagctgcgtc	tcatccaagg	ggtctgcgtc	cacggtaaag	accccgggca	gcaggcgcgc	6600
gtcgaagtag	tctatcttgc	atccttgcaa	gtctagcgcc	tgctgccatg	cgcgggcggc	6660
aagcgcgcgc	tcgtatgggt	tgagtggggg	accccatggc	atggggtggg	tgagcgcgga	6720
ggcgtacatg	ccgcaaatgt	cgtaaacgta	gaggggctct	ctgagtattc	caagatatgt	6780
agggtagcat	cttccaccgc	ggatgctggc	gcgcacgtaa	tcgtatagtt	cgtgcgaggg	6840
agcgaggagg	tcgggaccga	ggttgctacg	ggcgggctgc	tctgctcgga	agactatctg	6900
cctgaagatg	gcatgtgagt	tggatgatat	ggttggacgc	tggaagacgt	tgaagctggc	6960
gtctgtgaga	cctaccgcgt	cacgcacgaa	ggaggcgtag	gagtcgcgca	gcttgttgac	7020
cageteggeg	gtgacctgca	cgtctagggc	gcagtagtcc	agggtttcct	tgatgatgtc	7080
atacttatcc	tatccctttt	ttttccacag	ctcgcggttg	aggacaaact	cttcgcggtc	7140
tttccagtac	tcttggatcg	gaaacccgtc	ggcctccgaa	cggtaagagc	ctagcatgta	7200
gaactggttg	acggcctggt	aggcgcagca	tcccttttct	acgggtagcg	cgtatgcctg	7260
cacaacette	cggagcgagg	tgtgggtgag	cgcaaaggtg	tccctgacca	tgactttgag	7320
gtactggtat	ttgaagtcag	tgtcgtcgca	tccgccctgc	tcccagagca	aaaagtccgt	7380
acacttttta	gaacgcggat	ttggcagggc	gaaggtgaca	tcgttgaaga	gtatctttcc	7440
cococoadac	ataaaqttqc	gtgtgatgcg	gaagggtccc	ggcacctcgg	aacggttgtt	7500
aattacctgg	gcggcgagca	cgatctcgtc	aaagccgttg	atgttgtggc	ccacaatgta	7560
aagttccaag	aagcgcggga	tgcccttgat	ggaaggcaat	tttttaagtt	cctcgtaggt	7620
gagetettea	ggggagctga	gcccgtgctc	tgaaagggcc	cagtctgcaa	gatgagggtt	7680
ggaagggacg	aatgagctcc	acaggtcacg	ggccattagc	atttgcaggt	ggtcgcgaaa	7740
ggtcctaaac	taacaaccta	tggccatttt	ttctggggtg	atgcagtaga	aggtaagcgg	7800
atcttattcc	cagcggtccc	atccaaggtt	cgcggctagg	tctcgcgcgg	cagtcactag	7860
aggeteatet	ccgccgaact	tcatgaccag	catgaagggc	acgagctgct	tcccaaaggc	7920
ccccatccaa	gtataggtct	ctacatcgta	ggtgacaaag	agacgctcgg	tgcgaggatg	7980
cgagccgatc	gggaagaact	ggatctcccg	ccaccaattg	gaggagtggc	tattgatgtg	8040
gtgaaagtag	aagtccctgc	gacgggccga	acactcgtgc	tggcttttgt	aaaaacgtgc	8100
acagtactag	cagcagtaca	cgggctgtac	atcctgcacg	aggttgacct	gacgaccgcg	8160
cacaaggaag	cagagtggga	atttgagccc	ctcgcctggc	gggtttggct	ggtggtcttc	8220
tacttcggct	acttatectt	gaccgtctgg	ctgctcgagg	ggagttacgg	tggatcggac	8280
caccacqccq	cgcgagccca	aagtccagat	gtccgcgcgc	ggcggtcgga	gcttgatgac	8340
aacatcgcgc	agatgggagc	tgtccatggt	ctggagctcc	cgcggcgtca	ggtcaggcgg	8400

WO 03/031588

gagctcctgc	aggtttacct	cgcatagacg	ggtcagggcg	cgggctagat	ccaggtgata	8460
cctaatttcc	aggggctggt	tggtggcggc	gtcgatggct	tgcaagaggc	cgcatccccg	8520
cggcgcgact	acggtaccgc	gcggcgggcg	gtgggccgcg	ggggtgtcct	tggatgatgc	8580
atctaaaagc	ggtgacgcgg	gcgagccccc	ggaggtaggg	ggggctccgg	acccgccggg	8640
agagggggca	ggggcacgtc	ggcgccgcgc	gcgggcagga	gctggtgctg	cgcgcgtagg	8700
					ctgcgtgaag	8760
					ttcggtgtcg	8820
					ataggcgatc	8880
					tcgctccacg	8940
gtggcggcga	ggtcgttgga	aatgcgggcc	atgagctgcg	agaaggcgtt	gaggcctccc	9000
				cgcgggcgcg		9060
				agtttcgcag		9120
				agtacataac		9180
				ccatggcctc		9240
				ttaactcctc		9300
				aggctacagg		9360
tcttcttcaa	tctcctcttc	cataagggcc	tccccttctt	cttcttctgg	cggcggtggg	9420
				ggtcgacaaa		9480
atctccccgc	ggcgacggcg	catggtctcg	gtgacggcgc	ggccgttctc	gcgggggcgc	9540
				gcggggggct		9600
				taggtactcc		9660
gacctgagcg	agtccgcatc	gaccggatcg	gaaaacctct	cgagaaaggc	gtctaaccag	9720
tcacagtcgc	aaggtaggct	gagcaccgtg	gcgggcggca	gcgggcggcg	gtcggggttg	9780
tttctggcgg	aggtgctgct	gatgatgtaa	ttaaagtagg	cggtcttgag	acggcggatg	9840
				tgcgcaggcg		9900
ccccaggctt	cgttttgaca	tcggcgcagg	tctttgtagt	agtcttgcat	gagcctttct	9960
				ttgcatctat		10020
				tgcgtgtgac		10080
				cggctaatat		10140
				agcggtggta		10200
				cggtctggtg		10260
				caaatacgta		10320
				gcggctggcg		10380
				acataaggcg		10440
				tggaggcgcg		10500
cggacgcggt	tccagatgtt	gcgcagcggc	aaaaagtgct	ccatggtcgg	gacgctctgg	10560
				aaaaggagag		10620
				tcatggcgga		10680
tcgagccccg	tatccggccg	tccgccgtga	tccatgcggt	taccgcccgc	gtgtcgaacc	10740
caggtgtgcg	acgtcagaca	acgggggagt	gctccttttg	gcttccttcc	aggcgcggcg	10800
				gtaagcggtt		10860
				tattttccaa		10920
				gcgaacgggg		10980
ccgtcatgca	agaccccgct	tgcaaattcc	tccggaaaca	gggacgagcc	ccttttttgc	11040
ttttcccaga	tgcatccggt	gctgcggcag	atgcgccccc	ctcctcagca	gcggcaagag	11100
				ctaccgcgtc		11160
				ccccgcggcg		11220
				taggagcgcc		11280
				cgtacgtgcc		11340
				gggatcgaaa		11400
				tgcgcgagga		11460
				tggcggccgc		11520
				aaaaaagctt		11580
				tgatgcatct		11640
				tggcgcagct		11700

gtgcagcaca	gcagggacaa	cgaggcattc	agggatgcgc	tgctaaacat	agtagagccc	11760
gagggccgct	ggctgctcga	tttgataaac	atcctgcaga	gcatagtggt	gcaggagcgc	11820
agcttgagcc	tggctgacaa	ggtggccgcc	atcaactatt	ccatgcttag	cctgggcaag	11880
ttttacgccc	gcaagatata	ccatacccct	tacgttccca	tagacaagga	ggtaaagatc	11940
gaggggttct	acatgcgcat	ggcgctgaag	gtgcttacct	tgagcgacga	cctgggcgtt	12000
tatcgcaacg	agcgcatcca	caaggccgtg	agcgtgagcc	ggcggcgcga	gctcagcgac	12060
cgcgagctga	tgcacagcct	gcaaagggcc	ctggctggca	cgggcagcgg	cgatagagag	12120
gccgagtcct	actttgacgc	gggcgctgac	ctgcgctggg	ccccaagccg	acgcgccctg	12180
gaggcagctg	gggccggacc	tgggctggcg	gtggcacccg	cgcgcgctgg	caacgtcggc	12240
ggcgtggagg	aatatgacga	ggacgatgag	tacgagccag	aggacggcga	gtactaagcg	12300
gtgatgtttc	tgatcagatg	atgcaagacg	caacggaccc	ggcggtgcgg	geggegetge	12360
agagccagcc	gtccggcctt	aactccacgg	acgactggcg	ccaggtcatg	gaccgcatca	12420
tgtcgctgac	tgcgcgcaat	cctgacgcgt	tccggcagca	gccgcaggcc	aaccggctct	12480
ccgcaattct	ggaagcggtg	gtcccggcgc	gcgcaaaccc	cacgcacgag	aaggtgctgg	12540
cgatcgtaaa	cgcgctggcc	gaaaacaggg	ccatccggcc	cgacgaggcc	ggcctggtct	12600
acgacgcgct	gcttcagcgc	gtggctcgtt	acaacagcgg	caacgtgcag	accaacctgg	12660
accggctggt	gggggatgtg	cgcgaggccg	tggcgcagcg	tgagcgcgcg	cagcagcagg	12720
gcaacctggg	ctccatggtt	gcactaaacg	ccttcctgag	tacacagccc	gccaacgtgc	12780
cacaaaaaa	ggaggactac	accaactttg	tgagcgcact	gcggctaatg	gtgactgaga	12840
caccgcaaag	tgaggtgtac	cagtctgggc	cagactattt	tttccagacc	agtagacaag	12900
acctacagac	cgtaaacctg	agccaggctt	tcaaaaactt	gcaggggctg	tggggggtgc	12960
gggctcccac	aggcgaccgc	gcgaccgtgt	ctagcttgct	gacgcccaac	tcgcgcctgt	13020
tactactact	aatagcgccc	ttcacggaca	gtggcagcgt	gtcccgggac	acatacctag	13080
gtcacttgct	gacactgtac	cgcgaggcca	taggtcaggc	gcatgtggac	gagcatactt	13140
tccaggagat	tacaagtgtc	agccgcgcgc	tggggcagga	ggacacgggc	agcctggagg	13200
caaccctaaa	ctacctgctg	accaaccggc	ggcagaagat	cccctcgttg	cacagtttaa	13260
acagcgagga	ggagcgcatt	ttgcgctacg	tgcagcagag	cgtgagcctt	aacctgatgc	13320
gcgacggggt	aacgcccagc	gtggcgctgg	acatgaccgc	gcgcaacatg	gaaccgggca	13380
tgtatgcctc	aaaccggccg	tttatcaacc	gcctaatgga	ctacttgcat	cgcgcggccg	13440
ccgtgaaccc	cgagtatttc	accaatgcca	tcttgaaccc	gcactggcta	ccgccccctg	13500
gtttctacac	cgggggattc	gaggtgcccg	agggtaacga	tggattcctc	tgggacgaca	13560
tagacgacag	cgtgttttcc	ccgcaaccgc	agaccctgct	agagttgcaa	cagcgcgagc	13620
aggcagaggc	ggcgctgcga	aaggaaagct	tccgcaggcc	aagcagcttg	tccgatctag	13680
acactacaac	cccgcggtca	gatgctagta	gcccatttcc	aagcttgata	gggtctctta	13740
ccagcactcg	caccacccgc	ccgcgcctgc	tgggcgagga	ggagtaccta	aacaactcgc	13800
tgctgcagcc	gcagcgcgaa	aaaaacctgc	ctccggcatt	tcccaacaac	gggatagaga	13860
gcctagtgga	caagatgagt	agatggaaga	cgtacgcgca	ggagcacagg	gacgtgccag	13920
acccacaccc	gcccacccgt	cgtcaaaggc	acgaccgtca	gcggggtctg	gtgtgggagg	13980
acgatgactc	ggcagacgac	agcagcgtcc	tggatttggg	agggagtggc	aacccgtttg	14040
cocaccttco	ccccaggctg	gggagaatgt	tttaaaaaaa	aaaaagcatg	atgcaaaata	14100
aaaaactcac	caaggccatg	gcaccgagcg	ttggttttct	tgtattcccc	ttagtatgcg	14160
acacacaaca	atgtatgagg	aaggtcctcc	tccctcctac	gagagtgtgg	tgagcgcggc	14220
accagtageg	gcggcgctgg	gttctccctt	cgatgctccc	ctggacccgc	cgtttgtgcc	14280
tccgcggtac	ctgcggccta	ccggggggag	aaacagcatc	cgttactctg	agttggcacc	14340
cctattcgac	accacccgtg	tgtacctggt	ggacaacaag	tcaacggatg	tggcatccct	14400
gaactaccag	aacgaccaca	gcaactttct	gaccacggtc	attcaaaaca	atgactacag	14460
cccaaaaaaa	gcaagcacac	agaccatcaa	tcttgacgac	cggtcgcact	ggggcggcga	14520
cctgaaaacc	atcctgcata	ccaacatgcc	aaatgtgaac	gagttcatgt	ttaccaataa	14580
atttaaggcg	cgggtgatgg	tgtcgcgctt	gcctactaag	gacaatcagg	tggagctgaa	14640
atacgagtgg	gtggagttca	cgctgcccga	gggcaactac	tccgagacca	tgaccataga	14700
ccttatgaac	aacgcgatcg	tggagcacta	cttgaaagtg	ggcagacaga	acggggttct	14760
adayaacayc	atcggggtaa	agtttgacac	ccgcaacttc	agactggggt	ttgaccccgt	14820
cactootett	gtcatgcctg	gggtatatac	aaacgaagcc	ttccatccag	acatcatttt	14880
actaccaaca	tgcggggtgg	acttcaccca	cagecaceta	agcaacttgt	tgggcatccg	14940
2ccacca23a	cccttccagg	agggetttag	gatcacctac	gatgatctgg	agggtggtaa	15000
caageggeaa	222220039		→ = .			

						15060
cattcccgca	ctgttggatg	tggacgccta	ccaggcgagc	ttgaaagatg	acaccgaaca	15060
gggcgggggt	ggcgcaggcg	gcagcaacag	cagtggcagc	ggcgcggaag	agaactccaa	15120
cgcggcagcc	gcggcaatgc	agccggtgga	ggacatgaac	gatcatgcca	ttcgcggcga	15180
cacctttgcc	acacgggctg	aggagaagcg	cgctgaggcc	gaagcagcgg	ccgaagctgc	15240
cgcccccgct	gcgcaacccg	aggtcgagaa	gcctcagaag	aaaccggtga	tcaaacccct	15300
gacagaggac	agcaagaaac	gcagttacaa	cctaataagc	aatgacagca	ccttcaccca	15360
gtaccgcagc	tggtaccttg	catacaacta	cggcgaccct	cagaccggaa	tccgctcatg	15420
gaccctgctt	tgcactcctg	acgtaacctg	cggctcggag	caggtctact	ggtcgttgcc	15480
agacatgatg	caagaccccg	tgaccttccg	ctccacgcgc	cagatcagca	actttccggt	15540
ggtgggcgcc	gagctgttgc	ccgtgcactc	caagagcttc	tacaacgacc	aggccgtcta	15600
ctcccaactc	atccgccagt	ttacctctct	gacccacgtg	ttcaatcgct	ttcccgagaa	15660
ccagattttg	gcgcgcccgc	cagcccccac	catcaccacc	gtcagtgaaa	acgttcctgc	15720
tctcacagat	cacgggacgc	taccgctgcg	caacagcatc	ggaggagtcc	agcgagtgac	15780
cattactgac	gccagacgcc	gcacctgccc	ctacgtttac	aaggccctgg	gcatagtctc	15840
gccgcgcgtc	ctatcgagcc	gcactttttg	agcaagcatg	tccatcctta	tatcgcccag	15900
caataacaca	ggctggggcc	tgcgcttccc	aagcaagatg	tttggcgggg	ccaagaagcg	15960
ctccgaccaa	cacccagtgc	gcgtgcgcgg	gcactaccgc	gcgccctggg	gcgcgcacaa	16020
acgcggccgc	actgggcgca	ccaccgtcga	tgacgccatc	gacgcggtgg	tggaggaggc	16080
gcgcaactac	acgcccacgc	cgccaccagt	gtccacagtg	gacgcggcca	ttcagaccgt	16140
ggtgcgcgga	gcccggcgct	atgctaaaat	gaagagacgg	cggaggcgcg	tagcacgtcg	16200
ccaccgccgc	cgacccggca	ctgccgccca	acgcgcggcg	gcggccctgc	ttaaccgcgc	16260
acgtcgcacc	ggccgacggg	cggccatgcg	ggccgctcga	aggctggccg	cgggtattgt	16320
cactgtgccc	cccaggtcca	ggcgacgagc	ggccgccgca	gcagccgcgg	ccattagtgc	16380
tatgactcag	ggtcgcaggg	gcaacgtgta	ttgggtgcgc	gactcggtta	gcggcctgcg	16440
cgtgcccgtg	cgcacccgcc	ccccgcgcaa	ctagattgca	agaaaaaact	acttagactc	16500
gtactgttgt	atgtatccag	cggcggcggc	gcgcaacgaa	gctatgtcca	agcgcaaaat	16560
caaagaagag	atgctccagg	tcatcgcgcc	ggagatctat	ggccccccga	agaaggaaga	16620
gcaggattac	aagccccgaa	agctaaagcg	ggtcaaaaag	aaaaagaaag	atgatgatga	16680
tgaacttgac	gacgaggtgg	aactgctgca	cgctaccgcg	cccaggcgac	gggtacagtg	16740
gaaaggtcga	cgcgtaaaac	gtgttttgcg	acccggcacc	accgtagtct	ttacgcccgg	16800
tgagcgctcc	acccgcacct	acaagcgcgt	gtatgatgag	gtgtacggcg	acgaggacct	16860
gcttgagcag	gccaacgagc	gcctcgggga	gtttgcctac	ggaaagcggc	ataaggacat	16920
gctggcgttg	ccgctggacg	agggcaaccc	aacacctagc	ctaaagcccg	taacactgca	16980
gcaggtgctg	cccgcgcttg	caccgtccga	agaaaagcgc	ggcctaaagc	gcgagtctgg	17040
tgacttggca	cccaccgtgc	agctgatggt	acccaagcgc	cagcgactgg	aagatgtctt	17100
ggaaaaaatg	accgtggaac	ctgggctgga	gcccgaggtc	cgcgtgcggc	caatcaagca	17160
ggtggcgccg	ggactgggcg	tgcagaccgt	ggacgttcag	atacccacta	ccagtagcac	17220
cagtattgcc	accgccacag	agggcatgga	gacacaaacg	tccccggttg	cctcagcggt	17280
ggcggatgcc	gcggtgcagg	cggtcgctgc	ggccgcgtcc	aagacctcta	cggaggtgca	17340
aacggacccg	tggatgtttc	gcgtttcagc	ccccggcgc	ccgcgcggtt	cgaggaagta	17400
cggcgccgcc	agcgcgctac	tgcccgaata	tgccctacat	ccttccattg	cgcctacccc	17460
cggctatcgt	ggctacacct	accgccccag	aagacgagca	actacccgac	gccgaaccac	17520
cactggaacc	cgccgccgcc	gtcgccgtcg	ccagcccgtg	ctggccccga	tttccgtgcg	17580
cagggtggct	cgcgaaggag	gcaggaccct	ggtgctgcca	acagcgcgct	accaccccag	17640
catcgtttaa	aagccggtct	ttgtggttct	tgcagatatg	gccctcacct	gccgcctccg	17700
tttcccggtg	ccgggattcc	gaggaagaat	gcaccgtagg	aggggcatgg	ccggccacgg	17760
cctgacgggc	ggcatgcgtc	gtgcgcacca	ccggcggcgg	cgcgcgtcgc	accgtcgcat	17820
gcgcggcggt	atcetgeece	tccttattcc	actgatcgcc	gcggcgattg	gcgccgtgcc	17880
cogaattgca	tccgtggcct	tgcaggcgca	gagacactga	ttaaaaacaa	gttgcatgtg	17940
gaaaaatcaa	aataaaaagt	ctggactctc	acgctcgctt	ggtcctgtaa	ctattttgta	18000
gaatggaaga	catcaacttt	gcgtctctgg	ccccgcgaca	cggctcgcgc	ccgttcatgg	18060
gaaactggca	agatatcqqc	accagcaata	tgagcggtgg	cgccttcagc	tggggctcgc	18120
tgtggagcgg	cattaaaaat	ttcggttcca	ccgttaagaa	ctatggcagc	aaggcctgga	18180
acagcagcac	aggccagatg	ctgagggata	agttgaaaga	gcaaaatttc	caacaaaagg	18240
togtagatog	cctggcctct	ggcattagcq	gggtggtgga	cctggccaac	caggcagtgc	18300
3335	55					

	********	cttgatcccc	acceteceat	agaggagcct	ccaccaacca	18360
aaaataagat	taacagtaag	gggcgtggcg	aaaaacatcc	acacccasc	agggaagaaa	18420
tggagacagt	gcccccagag	gagcctccct	catacaaaaa	ggcactaaag	caaggcctgc	18480
ctctggtgac	gcaaatagac	cccatggcta	acacacatact	addecadese	acacccotaa	18540
ccaccacccg	teceategeg	cccatggcta	accacasacc	tatactacca	ggcccgaccg	18600
cgctggacct	geeteeeee	gccgacaccc	agcagaaacc	caccaccaac	ggtccgcgat	18660
ccgttgttgt	aacecgteet	agccgcgcgt	aaagagagt	gazcagcatc	ataaatetaa	18720
cgttgcggcc	cgtagccagt	ggcaactggc	tatanataa	taacatatca	tatotototo	18780
gggtgcaatc	cctgaagcgc	cgacgatgct	ctagtagacage	accacacac	cgctttccaa	18840
atgtatgcgt	ccatgtegee	gccagaggag	atatteata	cacatctcgg	accaggacgc	18900
gatggctacc	ccttcgatga	tgccgcagtg ggctggtgca	atttaccar	accaccaaga	cgtacttcag	18960
ctcggagtac	ctgagccccg	ggetggtgta	geeegeeege	caccaccatga	ccacagaccg	19020
cctgaataac	aagtttagaa	acccacggt	tataaaccat	gaggatactg	cotactcota	19080
gtcccagcgt	ttgacgctgc	ggttcatccc ctgtgggtga	taaccatata	crocacatoo	cttccacgta	19140
caaggcgcgg	ttcaccctag	tggacagggg	ccctactttt	aarccctact	ctagcactac	19200
ctttgacatc	egeggegege	cggacagggg	apatecttee	gaatgggatg	aagctgctac	19260
ctacaacgcc	etggeteeca	agggtgcccc	adatteetege	gaacgggacg	tagacgagca	19320
tgctcttgaa	ataaacctag	aagaagagga	cyatgataat	tattctccta	taaatattac	19380
agctgagcag	caaaaaactc	acgtatttgg	tonancect	aaatatocco	ataaaacatt	19440
aaaggagggt	attcaaatag	gtgtcgaagg	ataatacacc	actranatta	atcatgcagc	19500
tcaacctgaa	cctcaaatag	gagaatctca	gracyaa	tacqqttcat	atgraaaacc	19560
tgggagagtc	cttaaaaaga	ctaccccaat	tataaaaaa	caeggtttat	acctagaaag	19620
cacaaatgaa	aatggagggc	aaggcattct	tgtaaagtaa	gcagggaa	atastasctt	19680
tcaagtggaa	atgcaatttt	tctcaactac	tgaggegace	geaggeaatg	acactcatat	19740
gactcctaaa	gtggtattgt	acagtgaaga	cytagatata	gaaaccccag	aacaatctat	19800
ttcttacatg	cccactatta	aggaaggtaa	ccacyayaa	ottaatgggcc	tatattacaa	19860
gcccaacagg	cctaattaca	ttgcttttag	ggacaatttt	accygectaa	ctattataa	19920
cagcacgggt	aatatgggtg	ttctggcggg	ccaagcatcg	cayccyaacy	ctgttgtaga	19980
tttgcaagac	agaaacacag	agctttcata	ccagcility	tateatecae	ttggtgatag	20040
aaccaggtac	ttttctatgt	ggaatcaggc	tgttgacage	tacgacccag	tagatatat	20100
tattgaaaat	catggaactg	aagatgaact	tccaaattac	egettteeae	atagataga	20160
gattaataca	gagactctta	ccaaggtaaa	acctaaaaca	ggtcaggaaa	atggatggga	20220
aaaagatgct	acagaatttt	cagataaaaa	tgaaataaga	gttggaaata	tagageteta	20280
ggaaatcaat	ctaaatgcca	acctgtggag	aaatttcctg	tactccaaca	tagcgctgta	20340
tttgcccgac	aagctaaagt	acagtccttc	caacgtaaaa	atticigata	acccaaacac	20400
ctacgactac	atgaacaagc	gagtggtggc	tecegggtta	grggaerger	acattaacct	20460
tggagcacgc	tggtcccttg	actatatgga	caacgtcaac	ccatttaacc	accaccgcaa	20520
tgctggcctg	cgctaccgct	caatgttgct	gggcaatggt	cgctatgtgc	ccttccacat	20580
ccaggtgcct	cagaagttct	ttgccattaa	aaacctcctt	ctcctgccgg	gctcatacac	20640
ctacgagtgg	aacttcagga	aggatgttaa	catggttctg	cagagetece	taggaaatga	20700
cctaagggtt	gacggagcca	gcattaagtt	tgatagcatt	tgeetttaeg	ccaccttctt	20760
ccccatggcc	cacaacaccg	cctccacgct	tgaggccatg	cttagaaacg	acaccaacga	20700
ccagtccttt	aacgactatc	teteegeege	caacatgctc	taccctatac	ccgccaacgc	20820
taccaacgtg	cccatatcca	teceteeeg	caactgggcg	gctttccgcg	gctgggcctt	20940
cacgcgcctt	aagactaagg	aaaccccatc	actgggctcg	ggctacgacc	cttattacac	21000
ctactctggc	tctataccct	acctagatgg	aaccttttac	ctcaaccaca	cctttaagaa	21060
ggtggccatt	acctttgact	cttctgtcag	ctggcctggc	aatgaccgcc	tgcttacccc	21120
caacgagttt	gaaattaagc	gctcagttga	cggggagggt	tacaacgttg	cccagtgtaa	21120
catgaccaaa	gactggttcc	tggtacaaat	gctagctaac	tacaacattg	gctaccaggg	21180
cttctatatc	ccagagagct	acaaggaccg	catgtactcc	ttctttagaa	acttccagec	
catgageegt	caggtggtgg	atgatactaa	atacaaggac	taccaacagg	tgggcatect	21300
acaccaacac	aacaactctq	gatttgttgg	ctaccttgcc	cccaccatge	gcgaaggaca	21360 21420
gacctaccct	gctaacttcc	cctatccgct	tataggcaag	accgcagttg	acagcattac	
ccagaaaaag	tttctttgcg	atcgcaccct	ttggcgcatc	ccattctcca	gtaactttat	21480
gtccatgggc	gcactcacag	acctgggcca	aaaccttctc	tacgccaact	ccgcccacgc	21540
gctagacatg	acttttgagg	tggatcccat	ggacgagccc	accettett	atgttttgtt	21600

				ggcgtcatcg		21660
cctgcgcacg	cccttctcgg	ccggcaacgc	cacaacataa	agaagcaagc	aacatcaaca	21720
				ccattgtcaa		21780
				caggctttgt		21840
aagctcgcct	gcgccatagt	caatacggcc	ggtcgcgaga	ctgggggcgt	acactggatg	21900
gcctttgcct	ggaacccgca	ctcaaaaaca	tgctacctct	ttgagccctt	tggcttttct	21960
gaccagcgac	tcaagcaggt	ttaccagttt	gagtacgagt	cactcctgcg	ccgtagcgcc	22020
attgcttctt	ccccgaccg	ctgtataacg	ctggaaaagt	ccacccaaag	cgtacagggg	22080
cccaactcgg	ccgcctgtgg	actattctgc	tgcatgtttc	tccacgcctt	tgccaactgg	22140
ccccaaactc	ccatggatca	caaccccacc	atgaacctta	ttaccggggt	acccaactcc	22200
atgctcaaca	gtccccaggt	acagcccacc	ctgcgtcgca	accaggaaca	gctctacagc	22260
ttcctggagc	gccactcgcc	ctacttccgc	agccacagtg	cgcagattag	gagcgccact	22320
tctttttgtc	acttgaaaaa	catgtaaaaa	taatgtacta	gagacacttt	caataaaggc	22380
aaatgctttt	atttgtacac	tctcgggtga	ttatttaccc	ccacccttgc	cgtctgcgcc	22440
gtttaaaaat	caaaggggtt	ctgccgcgca	tegetatgeg	ccactggcag	ggacacgttg	22500
cgatactggt	gtttagtgct	ccacttaaac	tcaggcacaa	ccatccgcgg	cagctcggtg	22560
				ttagcaggtc		22620
atcttgaagt	cgcagttggg	gcctccgccc	tgcgcgcgcg	agttgcgata	cacagggttg	22680
cagcactgga	acactatcag	cgccgggtgg	tgcacgctgg	ccagcacgct	cttgtcggag	22740
atcagatccg	cgtccaggtc	ctccgcgttg	ctcagggcga	acggagtcaa	ctttggtagc	22800
				actcgcaccg		22860
aaaaggtgac	cgtgcccggt	ctgggcgtta	ggatacagcg	cctgcataaa	agccttgatc	22920
tgcttaaaag	ccacctgagc	ctttgcgcct	tcagagaaga	acatgccgca	agacttgccg	22980
gaaaactgat	tggccggaca	ggccgcgtcg	tgcacgcagc	accttgcgtc	ggtgttggag	23040
atctgcacca	catttcggcc	ccaccggttc	ttcacgatct	tggccttgct	agactgctcc	23100
				caatcacgtg		23160
				tctcagcgca		23220
				cctctgcaaa		23280
				tgttgctggt		23340
				cggccgccag		23400
				ccacgtggta		23460
				cgatcggcac		23520
				cctcttcctc		23580
ataccacgcg	ccactgggtc	gtcttcattc	agccgccgca	ctgtgcgctt	acctcctttg	23640
				tttgtagcgc		23700
ctttcttcct	cgctgtccac	gattacctct	ggtgatggcg	ggcgctcggg	cttgggagaa	23760
				ccgccgaggt		23820
				cttcctcgtc		23880
				gcggcggcga		23940
				cgcgtccgcg		24000
				cctataggca		24060
atggagtcag	tcgagaagaa	ggacagccta	accgccccct	ctgagttcgc	caccaccgcc	24120
				aggcaccccc		24180
				aagacgacga		24240
gtaccaacag	aggataaaaa	gcaagaccag	gacaacgcag	aggcaaacga	ggaacaagtc	24300
				gagacgacgt		24360
				aagagcgcag		24420
				tattctcacc		24480
				gcctcaactt		24540
				tccaaaactg		24600
				tggccttgcg		24660
				tctttgaggg		24720
				gcgaaaatga		24780
				ccgtactaaa		24840
gaggtcaccc	actttgccta	cccggcactt	aacctacccc	ccaaggtcat	gagcacagtc	24900
	-					

				_		24000
atgagtgagc	tgatcgtgcg	ccgtgcgcag	cccctggaga	gggatgcaaa	tttgcaagaa	24960
caaacagagg	agggcctacc	cgcagttggc	gacgagcagc	tagcgcgctg	gcttcaaacg	25020
cgcgagcctg	ccgacttgga	ggagcgacgc	aaactaatga	tggccgcagt	gctcgttacc	25080
gtggagcttg	agtgcatgca	gcggttcttt	gctgacccgg	agatgcagcg	caagctagag	25140
gaaacattgc	actacacctt	tcgacagggc	tacgtacgcc	aggcctgcaa	gatctccaac	25200
gtggagctct	gcaacctggt	ctcctacctt	ggaattttgc	acgaaaaccg	ccttgggcaa	25260
aacgtgcttc	attccacgct	caagggcgag	gcgcgccgcg	actacgtccg	cgactgcgtt	25320
tacttatttc	tatgctacac	ctggcagacg	gccatgggcg	tttggcagca	gtgcttggag	25380
gagtgcaacc	tcaaggagct	gcagaaactg	ctaaagcaaa	acttgaagga	cctatggacg	25440
accttcaacq	agcgctccgt	ggccgcgcac	ctggcggaca	tcattttccc	cgaacgcctg	25500
cttaaaaccc	tgcaacaggg	tctgccagac	ttcaccagtc	aaagcatgtt	gcagaacttt	25560
aggaacttta	tcctagagcg	ctcaggaatc	ttgcccgcca	cctgctgtgc	acttcctagc	25620
gactttgtgc	ccattaagta	ccgcgaatgc	cctccgccgc	tttggggcca	ctgctacctt	25680
ctgcagctag	ccaactacct	tgcctaccac	tctgacataa	tggaagacgt	gagcggtgac	25740
ggtctactgg	agtgtcactg	tcgctgcaac	ctatgcaccc	cgcaccgctc	cctggtttgc	25800
aattcgcagc	tgcttaacga	aagtcaaatt	atcggtacct	ttgagctgca	gggtccctcg	25860
cctgacgaaa	agtccgcggc	tccggggttg	aaactcactc	cggggctgtg	gacgtcggct	25920
taccttcgca	aatttgtacc	tgaggactac	cacgcccacg	agattaggtt	ctacgaagac	25980
caatcccqcc	cqccaaatgc	ggagcttacc	gcctgcgtca	ttacccaggg	ccacattctt	26040
ggccaattgc	aagccatcaa	caaagcccgc	caagagtttc	tgctacgaaa	gggacggggg	26100
gtttacttgg	acccccagtc	cggcgaggag	ctcaacccaa	tececegee	gccgcagccc	26160
tatcagcagc	agccgcgggc	ccttgcttcc	caggatggca	cccaaaaaga	agctgcagct	26220
gccgccgcca	cccacggacg	aggaggaata	ctgggacagt	caggcagagg	aggttttgga	26280
cgaggaggag	gaggacatga	tggaagactg	ggagagccta	gacgaggaag	cttccgaggt	26340
cgaagaggtg	tcagacgaaa	caccgtcacc	ctcggtcgca	ttcccctcgc	cggcgcccca	26400
gaaatcggca	accggttcca	gcatggctac	aacctccgct	cctcaggcgc	cgccggcact	26460
acccattcac	cgacccaacc	gtagatggga	caccactgga	accagggccg	gtaagtccaa	26520
gcagccgccg	ccgttagccc	aagagcaaca	acagcgccaa	ggctaccgct	catggcgcgg	26580
gcacaagaac	gccatagttg	cttgcttgca	agactgtggg	ggcaacatct	ccttcgcccg	26640
ccactttctt	ctctaccatc	acggcgtggc	cttcccccgt	aacatcctgc	attactaccg	26700
tcatctctac	agcccatact	gcaccggcgg	cagcggcagc	ggcagcaaca	gcagcggcca	26760
cacagaagca	aaggcgaccg	gatagcaaga	ctctgacaaa	gcccaagaaa	tccacagcgg	26820
cggcagcagc	aggaggagga	gcgctgcgtc	tggcgcccaa	cgaacccgta	tcgacccgcg	26880
agcttagaaa	caggattttt	cccactctgt	atgctatatt	tcaacagagc	aggggccaag	26940
aacaagagct	gaaaataaaa	aacaggtctc	tgcgatccct	cacccgcagc	tgcctgtatc	27000
acaaaaqcqa	agatcagctt	cggcgcacgc	tggaagacgc	ggaggctctc	ttcagtaaat	27060
actocococt	gactcttaag	gactagtttc	gcgccctttc	tcaaatttaa	gcgcgaaaac	27120
tacgtcatct	ccagcggcca	cacccggcgc	cagcacctgt	cgtcagcgcc	attatgagca	27180
aggaaattcc	cacqccctac	atgtggagtt	accagccaca	aatgggactt	gcggctggag	27240
ctgcccaaga	ctactcaacc	cgaataaact	acatgagcgc	gggaccccac	atgatatccc	27300
gggtcaacgg	aatccgcgcc	caccgaaacc	gaattctctt	ggaacaggcg	gctattacca	27360
ccacacctcq	taataacctt	aatccccgta	gttggcccgc	tgccctggtg	taccaggaaa	27420
atcccactcc	caccactgtg	gtacttccca	gagacgccca	ggccgaagtt	cagatgacta	27480
actcaggggc	gcagcttgcg	ggcggctttc	gtcacagggt	gcggtcgccc	gggcagggta	27540
taactcacct	gacaatcaga	gggcgaggta	ttcagctcaa	cgacgagtcg	gtgagctcct	27600
cacttaatct	ccqtccqqac	gggacatttc	agatcggcgg	cgccggccgt	ccttcattca	27660
cocctcotca	ggcaatccta	actctgcaga	cctcgtcctc	tgagccgcgc	tctggaggca	27720
ttggaactct	gcaatttatt	gaggagtttg	tgccatcggt	ctactttaac	cccttctcgg	27780
gacctcccqq	ccactatccg	gatcaattta	ttcctaactt	tgacgcggta	aaggactcgg	27840
cggacggcta	cgactgaatg	ttaagtggag	aggcagagca	actgcgcctg	aaacacctgg	27900
tccactatca	ccgccacaag	tgctttgccc	gcgactccgg	tgagttttgc	tactttgaat	27960
tacccaaqqa	tcatatcgag	ggcccggcgc	acggcgtccg	gcttaccgcc	cagggagagc	28020
ttacccataa	cctgattcgg	gagtttaccc	agcgccccct	gctagttgag	cgggacaggg	28080
gaccctgtgt	tctcactgtg	atttgcaact	gtcctaacct	tggattacat	caagatcttt	28140
gttgccatct	ctgtgctgag	tataataaat	acagaaatta	aaatatactg	gggctcctat	28200

cgccatcctg	taaacgccac	cgtcttcacc	cgcccaagca	aaccaaggcg	aaccttacct	28260
ggtactttta	acatctctcc	ctctgtgatt	tacaacagtt	tcaacccaga	cggagtgagt	28320
ctacgagaga	acctctccga	gctcagctac	tccatcagaa	aaaacaccac	cctccttacc	28380
tgccgggaac	gtacgagtgc	gtcaccggcc	gctgcaccac	acctaccgcc	tgaccgtaaa	28440
ccagactttt	tccggacaga	cctcaataac	tctgtttacc	agaacaggag	gtgagcttag	28500
aaaaccctta	gggtattagg	ccaaaggcgc	agctactgtg	gggtttatga	acaattcaag	28560
caactctacg	ggctattcta	attcaggttt	ctctagaatc	ggggttgggg	ttattctctg	28620
tcttqtqatt	ctctttattc	ttatactaac	gcttctctgc	ctaaggctcg	ccgcctgctg	28680
totocacatt	tgcatttatt	gtcagctttt	taaacgctgg	ggtcgccacc	caagatgatt	28740
aggtacataa	tcctaggttt	actcaccctt	gcgtcagccc	acggtaccac	ccaaaaggtg	28800
gattttaagg	agccagcctg	taatgttaca	ttcgcagctg	aagctaatga	gtgcaccact	28860
cttataaaat	gcaccacaga	acatgaaaag	ctgcttattc	gccacaaaaa	caaaattggc	28920
aagtatgctg	tttatgctat	ttggcagcca	ggtgacacta	cagagtataa	tgttacagtt	28980
ttccagggta	aaagtcataa	aacttttatg	tatacttttc	cattttatga	aatgtgcgac	29040
attaccatqt	acatgagcaa	acagtataag	ttgtggcccc	cacaaaattg	tgtggaaaac	29100
actggcactt	tctgctgcac	tgctatgcta	attacagtgc	tcgctttggt	ctgtacccta	29160
ctctatatta	aatacaaaag	cagacgcagc	tttattgagg	aaaagaaaat	gccttaattt	29220
actaaqttac	aaagctaatg	tcaccactaa	ctgctttact	cgctgcttgc	aaaacaaatt	29280
caaaaagtta	gcattataat	tagaatagga	tttaaacccc	ccggtcattt	cctgctcaat	29340
accattcccc	tgaacaattg	actctatgtg	ggatatgctc	cagcgctaca	accttgaagt	29400
caggetteet	ggatgtcagc	atctgacttt	ggccagcacc	tgtcccgcgg	atttgttcca	29460
gtccaactac	agcgacccac	cctaacagag	atgaccaaca	caaccaacgc	ggccgccgct	29520
accogactta	catctaccac	aaatacaccc	caagtttctg	cctttgtcaa	taactgggat	29580
aacttgggca	tataataatt	ctccatagcg	cttatgtttg	tatgccttat	tattatgtgg	29640
ctcatctgct	gcctaaagcg	caaacgcgcc	cgaccaccca	tctatagtcc	catcattgtg	29700
ctacacccaa	acaatgatgg	aatccataga	ttggacggac	tgaaacacat	gttcttttct	29760
cttacagtat	gattaaatga	gacatgattc	ctcgagtttt	tatattactg	acccttgttg	29820
cactttttta	tacatactec	acattggctg	cggtttctca	catcgaagta	gactgcattc	29880
cagcottcac	agtctatttg	ctttacggat	ttgtcaccct	cacgctcatc	tgcagcctca	29940
tcactgtggt	catcoccttt	atccagtgca	ttgactgggt	ctgtgtgcgc	tttgcatatc	30000
tcagacacca	tccccagtac	agggacagga	ctatagctga	gcttcttaga	attctttaat	30060
tatgaaattt	actotoactt	ttctgctgat	tatttqcacc	ctatctgcgt	tttgttcccc	30120
gacctccaag	cctcaaagac	atatatcatg	cagattcact	cgtatatgga	atattccaag	30180
ttoctacaat	gaaaaaagcg	atctttccga	agcctggtta	tatqcaatca	tctctgttat	30240
agtattetae	agtaccatct	tagccctagc	tatatatccc	taccttgaca	ttggctggaa	30300
acquataqat	accataaacc	acccaacttt	ccccacaccc	gctatgcttc	cactgcaaca	30360
agttattacc	gacaacttta	tcccagccaa	tcagcctcgc	cccacttctc	ccacccccac	30420
tgaaatcagc	tactttaatc	taacaggagg	agatgactga	caccctagat	ctagaaatgg	30480
acqqaattat	tacagagcag	cgcctgctag	aaaqacqcaq	ggcagcggcc	gagcaacagc	30540
gcatgaatca	agagetecaa	gacatggtta	acttgcacca	gtgcaaaagg	ggtatcttt	30600
gtctggtaaa	gcaggccaaa	gtcacctacg	acagtaatac	caccggacac	cgccttagct	30660
acaagttgcc	aaccaagcgt	cagaaattgg	tggtcatggt	gggagaaaag	cccattacca	30720
taactcagca	ctcggtagaa	accgaaggct	gcattcactc	accttgtcaa	ggacctgagg	30780
atctctgcac	ccttattaag	accctgtgcg	gtctcaaaga	tcttattccc	tttaactaat	30840
aaaaaaaaat	aataaagcat	cacttactta	aaatcaqtta	gcaaatttct	gtccagttta	30900
ttcagcagca	cctccttacc	ctcctcccag	ctctggtatt	gcagcttcct	cctggctgca	30960
aactttctcc	acaatctaaa	tggaatgtca	atttcctcct	attectatee	atccgcaccc	31020
actatettea	tattattaca	gatgaagcgc	gcaagaccgt	ctgaagatac	cttcaacccc	31080
gratatocat	atgacacgga	aaccggtcct	ccaactgtgc	cttttcttac	tectecettt	31140
gratecece	atgootttca	agagagtccc	cctggggtac	tctctttaca	cctatccgaa	31200
cctctagtta	cctccaatoo	catgcttgcg	ctcaaaatoo	gcaacggcct	ctctctggac	31260
gaggcggca	accttacctc	ccaaaatgta	accactataa	gcccacctct	caaaaaacc	31320
aantcaaaca	taaacctgga	aatatctgca	ccctcacag	ttacctcaga	agccctaact	31380
ataactacca	ccgcacctct	aatggtcgcg	ggcaacacac	tcaccatgca	atcacaggcc	31440
acaderacca	tacacaacte	caaacttagc	attgccaccc	aaggacccct	cacagtgtca	31500
cogocaacog	-30306					

				coaccatea	cartaccett	31560
gaaggaaagc	tagecetgea	aacatcagge	cccccacca	ccaccgatag	tracttrasa	31620
actatcactg	cctcacccc	tctaactact	gccactggta	gcttgggcat	tttacatata	31680
gagcccattt	atacacaaaa	tggaaaacta	ggactaaagt	acggggctcc	tastastact	31740
acagacgacc	taaacacttt	gaccgtagca	actygiccay	gtgtgactat	tatgcaactt	31800
tccttgcaaa	ctaaagttac	tggagccttg	ggccccgacc	cacaaggcaa	tatgeaucet	31860
aatgtagcag	gaggactaag	gattgattet	caaaacagac	gccttatact	tottttata	31920
tatccgtttg	atgctcaaaa	ccaactaaat	ctaagactag	gacagggccc	tacaccttca	31980
aactcagccc	acaacttgga	tattaactac	aacaaaggcc	tttacttgtt	atttaacact	32040
aacaattcca	aaaagcttga	ggttaaccta	ageactycea	aggggttgat	tacaccaaac	32100
acagccatag	ccattaatgc	aggagatggg	cttgaattty	gttcacctaa	caaccaaac	32160
acaaatcccc	tcaaaacaaa	aattggccat	ggcccagaac	ttgattcaaa	antannaaan	32220
gttcctaaac	taggaactgg	ccttagtttt	gacagcacag	gtgccattac	ctgtaggaaac	32280
aaaaataatg	ataagctaac	tttgtggacc	acaccagete	catctcctaa	tcaaatactt	32340
aatgcagaga	aagatgctaa	actcactttg	gtcttaacaa	aatgtggcag	aacagttcaa	32400
gctacagttt	cagttttggc	tgttaaaggc	agtitiggete	caatatctgg	ttccttccta	
agtgctcatc	ttattataag	atttgacgaa	aatggagtge	Lactadacaa	ttccttcctg	32520
gacccagaat	attggaactt	tagaaatgga	gatettacty	adgycacage	ctatacaaac	32580
gctgttggat	ttatgcctaa	cctatcagct	tatteaaaat	ctcacggtaa	aacegeeaaa	32640
agtaacattg	tcagtcaagt	ttacttaaac	ggagacaaaa	ctaaacctgt	ctctatctca	32700
attacactaa	acggtacaca	ggaaacagga	gacacaactc	caagtgcata	atcetettac	32760
ttttcatggg	actggtctgg	ccacaactac	accaatgaaa	tatttgccac	accecettat	32820
actttttcat	acattgccca	agaataaaga	account	ttatgtttca	caccaccaca	32880
ttttcaattg	cagaaaattt	caagtcattt	ttcattcagt	agtatagccc	traacctorr	32940
tagcttatac	agatcaccgt	accttaatca	aactcacaga	accctagtat	tasasarrat	33000
acctccctcc	caacacacag	agtacacagt	cetteetee	cggctggcct	catatageac	33060
catatcatgg	gtaacagaca	tattcttagg	tgttatatte	cacacggttt	teatateact	33120
caaacgctca	tcagtgatat	taataaactc	cccgggcagc	tcacttaagt	acadedsaga	33180
gtccagctgc	tgagccacag	gctgctgtcc	aacttgcggt	tgcttaacgg	geggegaagg	33240
agaagtccac	gcctacatgg	gggtagagtc	ataatcgtgc	atcaggatag	ggcggcggcg	33300
ctgcagcagc	gcgcgaataa	actgctgccg	cegeegetee	gtcctgcagg	ttatactacac	33360
ggcagtggtc	tcctcagcga	tgattcgcac	egeeegeage	ataaggcgcc	acaccaccac	33420
ggcacagcag	cgcaccctga	tctcacttaa	accagcacag	taactgcagc	caagcaccac	33480
aatattgttc	aaaatcccac	agtgcaaggc	getgtateta	aagctcatgg	ccctcataaa	33540
agaacccacg	tggccatcat	accacaagcg	caygrayarr	aagtggcgac	ccccatacca	33600
cacgctggac	ataaacatta	cctcttttgg	catguigua	ctaaccaccc	cccggtacca	33660
tataaacctc	tgattaaaca	tggcgccatc	caccaccatc	ctaaaccagc	rgaceaaaae	33720
ctgcccgccg	gctatacact	gcagggaacc	gggactggaa	atattaacac	ggagagccca	33780
ggactcgtaa	ccatggatca	tcatgctcgt	catgatatea	atgreggeae	aacacaggca	33840
cacgtgcata	cacttcctca	ggattacaag	ecceccege	grayaacca	tatcccaggg	33900
aacaacccat	tcctgaatca	gcgtaaatcc	cacactycay	ggaagacctc	ccantatoot	33960
cacgttgtgc	attgtcaaag	tgttacattc	gggcagcagc	ggatgatcct	tacaccaaaa	34020
agcgcgggtt	tetgteteaa	aaggaggtag	acgatetta	ecaccadaca	tgcgccgaga tagtcatatt	34080
caaccgagat	cgtgttggtc	gragratear	gccaaatgga	acgccggacg	tetegeeget	34140
tcctgaagca	aaaccaggtg	egggegtgae	togactet	gcgtctccgg	anacacccc	34200
tagatcgctc	tgtgtagtag	ttgtagtata	gegegetee	caaagcatcc	treaceaceg	34260
tggcttcggg	ttctatgtaa	actecticat	gegeegeege	casatcacac	tccaccaccg acgggaggag	34320
cagaataagc	cacacccagc	caacctacac	tttattcca	aaagattato	caaaacctca	34380
cgggaagagc	tggaagaacc	atguttutt	cetecatica	cataatcaaa	caaaacctca	34440
aaatgaagat	ctattaagtg	aacgegetee	tacacaataa	cttccaaaa	ctctacagcc	34500
aaagaacaga	taatggcatt	tgtaagatgt	catterage	gasteteete	gcaaacggcc	34560
ctcacgtcca	agtggacgta	aagyctaaac	teatetegg	accttctc22	tataaacatt tatatctcta	34620
ccagcacctt	caaccatgcc	caactaattc	ataaaaatat	actroagage	tatateteta	34680
agcaaatccc	gaatattaag	ccggccatt	guaadaauuu	tteeteeee	gccctccacc	34740
ttcagcctca	agcagcgaat	catgattgca	addattcayy	taggtccctt	acctgtataa	34800
gattcaaaag	cggaacatta	acaaaaatac	dycyateceg	Laggictic	cgcagggcca	

```
34860
gctgaacata atcgtgcagg tctgcacgga ccagcgcggc cacttccccg ccaggaacct
tgacaaaaga acccacactg attatgacac gcatactcgg agctatgcta accagcgtag
                                                                     34920
ccccgatgta agctttgttg catgggcggc gatataaaat gcaaggtgct gctcaaaaaa
                                                                     34980
tcaggcaaag cctcgcgcaa aaaagaaagc acatcgtagt catgctcatg cagataaagg
                                                                     35040
caggtaagct ccggaaccac cacagaaaaa gacaccattt ttctctcaaa catgtctgcg
                                                                     35100
ggtttctgca taaacacaaa ataaaataac aaaaaaacat ttaaacatta gaagcctqtc
                                                                     35160
                                                                     35220
ttacaacagg aaaaacaacc cttataagca taagacggac tacggccatg ccggcgtgac
                                                                     35280
cqtaaaaaaa ctqqtcaccg tgattaaaaa gcaccaccga cagctcctcg gtcatgtccg
gagtcataat gtaagactcg gtaaacacat caggttgatt catcggtcag tgctaaaaaag
                                                                    35340
cgaccgaaat agcccggggg aatacatacc cgcaggcgta gagacaacat tacagccccc
                                                                    35400
ataggaggta taacaaaatt aataggagag aaaaacacat aaacacctga aaaaccctcc
                                                                    35460
tgcctaggca aaatagcacc ctcccgctcc agaacaacat acagcgcttc acagcggcag
                                                                    35520
cctaacagtc agccttacca gtaaaaaaga aaacctatta aaaaaacacc actcgacacg
                                                                    35580
gcaccaqctc aatcagtcac agtgtaaaaa agggccaagt gcagagcgag tatatatagg
                                                                    35640
                                                                    35700
actaaaaaat gacgtaacgg ttaaagtcca caaaaaacac ccagaaaacc gcacgcgaac
                                                                    35760
ctacqcccaq aaacqaaaqc caaaaaaccc acaacttcct caaatcgtca cttccgtttt
                                                                    35820
cccacgttac gtaacttccc attttaagaa aactacaatt cccaacacat acaagttact
                                                                    35880
ccgccctaaa acctacgtca cccgccccgt tcccacgccc cgcgccacgt cacaaactcc
acccctcat tatcatattg gcttcaatcc aaaataaggt atattattga tgatg
                                                                    35935
<210> 9
<211> 35935
<212> DNA
<213> Adenovirus serotype 5
<400> 9
                                                                       60
catcatcaat aatatacctt attttggatt gaagccaata tgataatgag ggggtggagt
                                                                      120
ttgtgacgtg gcgcggggcg tgggaacggg gcgggtgacg tagtagtgtg gcggaagtgt
                                                                      180
qatgttgcaa gtgtggcgga acacatgtaa gcgacggatg tggcaaaagt gacgtttttg
                                                                      240
gtgtgcgccg gtgtacacag gaagtgacaa ttttcgcgcg gttttaggcg gatgttgtag
taaatttggg cgtaaccgag taagatttgg ccattttcgc gggaaaactg aataagagga
                                                                      300
                                                                      360
agtgaaatct gaataatttt gtgttactca tagcgcgtaa tatttgtcta gggccgcggg
                                                                      420
gactttgacc gtttacgtgg agactcgccc aggtgttttt ctcaggtgtt ttccgcgttc
                                                                      480
cgggtcaaag ttggcgtttt attattatag tcagctgacg tgtagtgtat ttatacccgg
tgagttcctc aagaggccac tcttgagtgc cagcgagtag agttttctcc tccgagccgc
                                                                      540
                                                                      600
tccgacaccg ggactgaaaa tgagacatat tatctgccac ggaggtgtta ttaccgaaga
                                                                      660
aatggccgcc agtcttttgg accagctgat cgaagaggta ctggctgata atcttccacc
tcctagccat tttgaaccac ctacccttca cgaactgtat gatttagacg tgacggcccc
                                                                      720
cgaagatccc aacgaggagg cggtttcgca gatttttccc gactctgtaa tgttggcggt
                                                                      780
gcaggaaggg attgacttac tcacttttcc gccggcgccc ggttctccgg agccgcctca
                                                                      840
                                                                      900
cctttcccgg cagcccgagc agccggagca gagagccttg ggtccggttt ctatgccaaa
                                                                      960
ccttgtaccg gaggtgatcg atcttacctg ccacgaggct ggctttccac ccagtgacga
                                                                     1020
cgaggatgaa gagggtgagg agtttgtgtt agattatgtg gagcaccccg ggcacggttg
                                                                     1080
caggicity cattatcacc ggaggaatac gggggaccca gatattatgt gticgctity
ctatatgagg acctgtggca tgtttgtcta cagtaagtga aaattatggg cagtgggtga
                                                                     1140
tagagtggtg ggtttggtgt ggtaattttt tttttaattt ttacagtttt gtggtttaaa
                                                                     1200
gaattttgta ttgtgatttt tttaaaaggt cctgtgtctg aacctgagcc tgagcccgag
                                                                     1260
ccagaaccgg agcctgcaag acctacccgc cgtcctaaaa tggcgcctgc tatcctgaga
                                                                     1320
cgcccgacat cacctgtgtc tagagaatgc aatagtagta cggatagctg tgactccggt
                                                                     1380
                                                                     1440
cettetaaca caceteetga gatacaceeg gtggteeege tgtgeeecat taaaceagtt
                                                                     1500
gccgtgagag ttggtgggcg tcgccaggct gtggaatgta tcgaggactt gcttaacgag
                                                                     1560
cctgggcaac ctttggactt gagctgtaaa cgccccaggc cataaggtgt aaacctgtga
ttgcgtgtgt ggttaacgcc tttgtttgct gaatgagttg atgtaagttt aataaagggt
                                                                     1620
                                                                     1680
gagataatgt ttaacttgca tggcgtgtta aatggggcgg ggcttaaagg gtatataatg
                                                                     1740
cgccgtgggc taatcttggt tacatctgac ctcatggagg cttgggagtg tttggaagat
```

	tgcgtaactt	actagaecan	acctctaaca	gracetettg	gttttggagg	1800
ttttctgctg	gctcatccca	gccggaacag	atctacagaa	traaggagga	ttacaagtgg	1860
tttctgtggg	agcttttgaa	atestateat	georgeagua	attetttgaa	tetaggteac	1920
gaatttgaag	tccaagagaa	accergege	actttggatt	tttccacacc	agagagact	1980
caggcgcttt	tccaagagaa	ggtcattaag	acceeggace	adadcasaas	aacccatcta	2040
gcggctgctg	ttgctttttt	gagicitaca	atacatatat	ggagegaaga	tataagacac	2100
agcggggggt	acctgctgga	-tetterage	caccatacae	taatacccac	adaddadcad	2160
aagaatcgcc	tgctactgtt	greereegre	cgcccggcga	agaggggatg	даасссаада	2220
cagcagcagc	aggaggaagc accctcggga	caggeggegg	tacacctcc	tgaactgtat	ccagaactga	2280
geeggeetgg	gacaattaca	acquatgecg	annanctasa	agaaataaaa	aggaagcagg	2340
gacgcatttt	gacaattata ggctacagag	gaggargggc	atctacettt	tagettaatg	accagacacc	2400
gggcttgtga	ggctacagag	gaggctagga	acceagette	cactaataaa	cttgatctgc	2460
gtcctgagtg	tattactttt	caacagatca	ccacttactq	actacaacca	ggggatgatt	2520
tggcgcagaa	gtattccata	tataaaaaaa	tageacttag	gccagattgc	aagtacaaga	2580
ttgaggaggc	tattagggta tgtaaatatc	tatgcaaagg	cotacetta	taaaacaaa	accasaataa	2640
tcagcaaact	tgtaaatate	aygaattgtt	getatacet	cataaatato	taaccaaaaa	2700
agatagatac	ggaggatagg	grygeerra	atataaaatt	tactggcccc	aattttagcg	2760
tgcttggcat	ggacggggtg	gitattatya	tagtaggee	tataaacttc	tatogottta	2820
gtacggtttt	cctggccaat	accaaccita	taaggattca	agactatacc	ttttactact	2880
acaatacctg	tgtggaagcc	tggaccgatg	gaagggttg	aattaaraaa	tacctcttta	2940
gctggaaggg	ggtggtgtgt	cgccccaaaa	geagggeeee	gatacaccac	aatgtggcct	3000
aaaggtgtac	cttgggtatc	ctgtctgagg	graderecay	ggtgtgttat	aacgtggcct	3060
ccgactgtgg	ttgcttcatg	ctagtgaaaa	gegtggetgt	gattaagtat	aactotcacc	3120
gtggcaactg	cgaggacagg	gcctctcaga	tyctyactty	ctcggacggc	tttgaggata	3180
tgctgaagac	cattcacgta	gccagccact	ctcgcaaggc	ccggccagcg	ctaccttacc	3240
acatactgac	ccgctgttcc	ttgcatttgg	gtaacaggag	gggggtgttc	angetenace	3300
aatgcaattt	gagtcacact	aagatattgc	ttgagcccga	gagcatgtcc	aaggrgaacc	3360
tgaacggggt	gtttgacatg	accatgaaga	tctggaaggt	getgaggtae	gatgagattt	3420
gcaccaggtg	cagaccctgc	gagtgtggcg	gtaaacatat	caggaaccag	cccgcgacgc	3480
tggatgtgac	cgaggagctg	aggcccgatc	acttggtgct	ggeetgeace	egegetgage	3540
ttggctctag	cgatgaagat	acagattgag	gtactgaaat	gtgtgggcgt	ggcttaaggg	3600
tgggaaagaa	tatataaggt	gggggtctta	tgtagttttg	tatctgtttt	geageageeg	3660
ccgccgccat	gagcaccaac	tcgtttgatg	gaagcattgt	gageteatat	retartage	3720
gcatgccccc	atgggccggg	gtgcgtcaga	atgtgatggg	ctecageatt	gatggttgtt	3780
ccgtcctgcc	cgcaaactct	actaccttga	cctacgagac	cgtgtctgga	acgccgttgg	3840
agactgcagc	ctccgccgcc	gcttcagccg	ctgcagccac	egecegeggg	attgtgactg	3900
actttgcttt	cctgagcccg	cttgcaagca	gtgcagcttc	ccgttcatcc	gcccgcgatg	3960
acaagttgac	ggctcttttg	gcacaattgg	attetttgac	ccgggaactt	aatgtegttt	4020
ctcagcagct	gttggatctg	cgccagcagg	tttctgccct	gaaggettee	tecectecca	4080
atgcggttta	aaacataaat	aaaaaaccag	actetgtttg	gatttggatt	aagcaagtgt	4140
cttgctgtct	ttatttaggg	gttttgcgcg	cgcggtaggc	ccgggaccag	eggicleggi	4200
cgttgagggt	cctgtgtatt	, ttttccagga	cgtggtaaag	gtgactctgg	atgttcagat	4260
acatgggcat	aagcccgtct	ctggggtgga	ggtagcacca	ctgcagagct	tcatgctgcg	4320
gggtggtgtt	gtagatgatc	cagtcgtagc	aggagcgctg	ggcgtggtgc	ctaaaaacgt	4320
ctttcagtag	caagctgatt	qccaggggca	ggcccttggt	gtaagtgttt	acaaageggu	
taagetggga	tagatacata	cgtggggata	tgagatgcat	cttggactgt	atttttaggt	4440 4500
tooctatott	cccagccata	tccctccggg	gattcatgtt	gtgcagaacc	accagcacag	4560
totatecoot	gcacttggga	aatttgtcat	gtagcttaga	aggaaatgcg	tggaagaact	
tagagacacc	cttgtgacct	ccaagatttt	ccatgcattc	gtccataatg	atggcaatgg	4620
acccacaaac	gacagectag	gcgaagatat	ttctgggatc	actaacgtca	tagttgtgtt	4680
ccaggatgag	atcotcatag	gccattttta	caaagcgcgg	gcggagggtg	ccagactgcg	4740
gtataatggt	tccatccggc	ccaggggcgt	agttaccctc	acagatttgc	atttcccacg	4800
criticactic	agatggggg	atcatgtcta	cctgcggggc	gatgaagaaa	acggtttccg	4860
gggtaggga	gatcagctgg	gaagaaagca	ggttcctgag	cagctgcgac	ttaccgcagc	4920
caataaaccc	gtaaatcaca	cctattaccg	ggtgcaactg	gtagttaaga	gagetgeage	4980
tgccgtcatc	cctgagcagg	ggggccactt	cgttaagcat	gtccctgact	cgcatgtttt	5040

ccctgaccaa	atccgccaga	aggcgctcgc	cgcccagcga	tagcagttct	tgcaaggaag	5100
caaagttttt	caacggtttg	agaccgtccg	ccgtaggcat	gcttttgagc	gtttgaccaa	5160
gcagttccag	gcggtcccac	agctcggtca	cctgctctac	ggcatctcga	tccagcatat	5220
ctcctcgttt	cgcgggttgg	ggcggctttc	gctgtacggc	agtagtcggt	gctcgtccag	5280
acgggccagg	gtcatgtctt	tccacgggcg	cagggtcctc	gtcagcgtag	tctgggtcac	5340
ggtgaagggg	tgcgctccgg	gctgcgcgct	ggccagggtg	cgcttgaggc	tggtcctgct	5400
ggtgctgaag	cgctgccggt	cttcgccctg	cgcgtcggcc	aggtagcatt	tgaccatggt	5460
gtcatagtcc	agcccctccg	cggcgtggcc	cttggcgcgc	agcttgccct	tggaggaggc	5520
				ttgggcgcga		5580
				gtctcgcatt		5640
				ccatgctttt		5700
				acgaaaaggc		5760
				ccgcggtcct		5820
				agcacgaagg		5880
ggaggggtag	cggtcgttgt	ccactagggg	gtccactcgc	tccagggtgt	gaagacacat	5940
gtcgccctct	tcggcatcaa	ggaaggtgat	tggtttgtag	gtgtaggcca	cgtgaccggg	6000
				tegtecteac		6060
atcgctgtct	gcgagggcca	gctgttgggg	tgagtactcc	ctctgaaaag	cgggcatgac	6120
ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	ttgatattca	cctggcccgc	6180
ggtgatgcct	ttgagggtgg	ccgcatccat	ctggtcagaa	aagacaatct	ttttgttgtc	6240
aagcttggtg	gcaaacgacc	cgtagagggc	gttggacagc	aacttggcga	tggagcgcag	6300
ggtttggttt	ttgtcgcgat	cggcgcgctc	cttggccgcg	atgtttagct	gcacgtattc	6360
				tcgtcgggca		6420
				gtggctacct		6480
gcgctcgttg	gtccagcaga	ggcggccgcc	cttgcgcgag	cagaatggcg	gtagggggtc	6540
tagctgcgtc	tcgtccgggg	ggtctgcgtc	cacggtaaag	accccgggca	gcaggcgcgc	6600
gtcgaagtag	tctatcttgc	atccttgcaa	gtctagcgcc	tgctgccatg	cgcgggcggc	6660
				atggggtggg		6720
ggcgtacatg	ccgcaaatgt	cgtaaacgta	gaggggctct	ctgagtattc	caagatatgt	6780
agggtagcat	cttccaccgc	ggatgctggc	gcgcacgtaa	tcgtatagtt	cgtgcgaggg	6840
agcgaggagg	tcgggaccga	ggttgctacg	ggcgggctgc	tctgctcgga	agactatctg	6900
				tggaagacgt		6960
				gagtcgcgca		7020
				agggtttcct		7080
				aggacaaact		7140
				cggtaagagc		7200
				acgggtagcg		7260
				tccctgacca		7320
				tcccagagca		7380
				tcgttgaaga		7440
				ggcacctcgg		7500
aattacctgg	gcggcgagca	cgatctcgtc	aaagccgttg	atgttgtggc	ccacaatgta	7560
aagttccaag	aagcgcggga	tgcccttgat	ggaaggcaat	tttttaagtt	cctcgtaggt	7620
				cagtctgcaa		7680
				atttgcaggt		7740
				atgcagtaga		7800
				tctcgcgcgg		7860
				acgagctgct		7920
				agacgctcgg		7980
				gaggagtggc		8040
				tggcttttgt		8100
				aggttgacct		8160
				gggtttggct		8220
				ggagttacgg		8280
caccacgccg	cgcgagccca	aagtccagat	gtccgcgcgc	ggcggtcgga	gcttgatgac	8340

				_		
aacatcgcgc	agatgggagc	tgtccatggt	ctggagctcc	cgcggcgtca	ggtcaggcgg	8400
gagctcctgc	aggtttacct	cgcatagacg	ggtcagggcg	cgggctagat	ccaggtgata	8460
cctaatttcc	aggggctggt	tggtggcggc	gtcgatggct	tgcaagaggc	cgcatccccg	8520
cggcgcgact	acggtaccgc	gcggcgggcg	gtgggccgcg	ggggtgtcct	tggatgatgc	8580
atctaaaagc	ggtgacgcgg	gcgagccccc	ggaggtaggg	ggggctccgg	acccgccggg	8640
agagggggca	ggggcacgtc	ggcgccgcgc	gcgggcagga	gctggtgctg	cgcgcgtagg	8700
ttgctggcga	acgcgacgac	gcggcggttg	atctcctgaa	tctggcgcct	ctgcgtgaag	8760
acgacgggcc	cggtgagctt	gagcctgaaa	gagagttcga	cagaatcaat	ttcggtgtcg	8820
ttgacggcgg	cctggcgcaa	aatctcctgc	acgtctcctg	agttgtcttg	ataggcgatc	8880
tcggccatga	actgctcgat	ctcttcctcc	tggagatctc	cgcgtccggc	tegetecacg	8940
				agaaggcgtt		9000
tcattccaga	cacaactata	gaccacgccc	ccttcggcat	cgcgggcgcg	catgaccacc	9060
tacacaaaat	tgaggtcgag	ataccagaca	aagacggcgt	agtttcgcag	gcgctgaaag	9120
aggragata	agataataac	gatatattet	accacgaaga	agtacataac	ccagcgtcgc	9180
				ccatggcctc		9240
accordant	trasasactr	ggagttgcgc	accascscaa	ttaactcctc	ctccagaaga	9300
acggcgaagt	caacaacaat	atcacacac	tracactoss	aggctacagg	gacetettet	9360
tattattan	tetestette	cataagggg	teceettett	cttcttctgg	caacaataaa	9420
				ggtcgacaaa		9480
ggagggggga	cacggeggeg	acyacygcyc	accaggagge	ggccgacaaa	acadadacac	9540
ateteeege	ggegaeggeg	catggeeteg	ttataaatta	ggccgttctc	accatacaac	9600
agttggaaga	egeegeeege	catgictegg	ccacgggccg	gcggggggct	gecaegege	9660
agggatacgg	cgctaacgat	gcatettaat	aactgctgtg	taggtactcc	gccgccgagg	9720
gacctgagcg	agtccgcatc	gaccggatcg	gaaaaccccc	cgagaaaggc	gtetaaccay	9780
tcacagtcgc	aaggtaggct	gagcaccgtg	gcgggcggca	gcgggcggcg	gccggggccg	
				cggtcttgag		9840
gtcgacagaa	gcaccatgtc	cttgggtccg	gcctgctgaa	tgcgcaggcg	greggeearg	9900
ccccaggctt	cgttttgaca	teggegeagg	tctttgtagt	agtcttgcat	gagcctttct	9960
accggcactt	cttcttctcc	ttcctcttgt	cctgcatctc	ttgcatctat	cgctgcggcg	10020
gcggcggagt	ttggccgtag	gtggcgccct	cttcctccca	tgcgtgtgac	cccgaagccc	10080
ctcatcggct	gaagcagggc	taggtcggcg	acaacgcgct	cggctaatat	ggcctgctgc	10140
acctgcgtga	gggtagactg	gaagtcatcc	atgtccacaa	agcggtggta	tgcgcccgtg	10200
ttgatggtgt	aagtgcagtt	ggccataacg	gaccagttaa	cggtctggtg	acccggctgc	10260
gagagctcgg	tgtacctgag	acgcgagtaa	gccctcgagt	caaatacgta	gtcgttgcaa	10320
gtccgcacca	ggtactggta	tcccaccaaa	aagtgcggcg	gcggctggcg	gtagaggggc	10380
cagcgtaggg	tggccggggc	tccgggggcg	agatcttcca	acataaggcg	atgatatccg	10440
tagatgtacc	tggacatcca	ggtgatgccg	gcggcggtgg	tggaggcgcg	cggaaagtcg	10500
cggacgcggt	tccagatgtt	gcgcagcggc	aaaaagtgct	ccatggtcgg	gacgctctgg	10560
ccaatcaaac	gcgcgcaatc	gttgacgctc	tagaccgtgc	aaaaggagag	cctgtaagcg	10620
ggcactcttc	cataatctaa	tggataaatt	cgcaagggta	tcatggcgga	cgaccggggt	10680
tcgagccccg	tatccggccg	tccgccgtga	tccatgcggt	taccgcccgc	gtgtcgaacc	10740
caggtgtgcg	acqtcaqaca	acqqqqqaqt	gctccttttg	gcttccttcc	aggcgcggcg	10800
actactacac	tagctttttt	ggccactggc	cgcgcgcagc	gtaagcggtt	aggctggaaa	10860
gcgaaagcat	taagtggctc	gctccctgta	gccggagggt	tattttccaa	gggttgagtc	10920
acaaaacccc	coattcaaat	ctcggaccgg	ccggactgcg	gcgaacgggg	gtttgcctcc	10980
ccatcataca	agaccccgct	tocaaattcc	tccggaaaca	gggacgagcc	ccttttttgc	11040
ttttcccaga	tgcatccggt	gctgcggcag	atgcgccccc	ctcctcagca	gcggcaagag	11100
caadacage	ggcagacatg	Caddacaccc	tecetecte	ctaccgcgtc	aggaggggg	11160
acateceee	ttaacacaac	aggatagt	gattacgaac	cccgcggcg	ccaaacccaa	11220
cactactca	acttonage	-20-20-330	ctaacacaac	taggagcgcc	ctctcctaaa	11280
cactaccegy	anataceast	daadcataat	acacatasaa	cgtacgtgcc	acaacaaaac	11340
ctatttaaa	aggracaget	adaddaddd	daddadatec	gggatcgaaa	gttccacgca	11400
aggaggaga	tacaacetaa	cctdaatccc	nadcoutter.	tgcgcgagga	ggactttgag	11460
gggcgcgage	raaccorrate	tantecene	202033cc3c	tggcggccgc	cgacctggta	11520
acceptage	ancanach	raaccarrar	attaactttc	aaaaaagctt	taacaaccac	11580
accycatacy	ttataacac	canacanta	actategree	tgatgcatct	atagaacttt	11640
gracacac	regragegeg	cyayyayyry	gulalayyac	cyacycatci	309994000	

			aagccgctca			11700
			agggatgcgc			11760
			atcctgcaga			11820
			atcaactatt			11880
			tacgttccca			11940
			gtgcttacct			12000
			agcgtgagcc			12060
cgcgagctga	tgcacagcct	gcaaagggcc	ctggctggca	cgggcagcgg	cgatagagag	12120
			ctgcgctggg			12180
			gtggcacccg			12240
			tacgagccag			12300
			caacggaccc			12360
			acgactggcg			12420
			tccggcagca			12480
			gcgcaaaccc			12540
cgatcgtaaa	cgcgctggcc	gaaaacaggg	ccatccggcc	cgacgaggcc	ggcctggtct	12600
			acaacagcgg			12660
accggctggt	gggggatgtg	cgcgaggccg	tggcgcagcg	tgagcgcgcg.	cagcagcagg	12720
gcaacctggg	ctccatggtt	gcactaaacg	ccttcctgag	tacacagccc	gccaacgtgc	12780
			tgagcgcact			12840
caccgcaaag	tgaggtgtac	cagtctgggc	cagactattt	tttccagacc	agtagacaag	12900
			tcaaaaactt			12960
			ctagcttgct			13020
			gtggcagcgt			13080
			taggtcaggc			13140
			tggggcagga			13200
caaccctaaa	ctacctgctg	accaaccggc	ggcagaagat	cccctcgttg	cacagtttaa	13260
acagcgagga	ggagcgcatt	ttgcgctacg	tgcagcagag	cgtgagcctt	aacctgatgc	13320
gcgacggggt	aacgcccagc	gtggcgctgg	acatgaccgc	gcgcaacatg	gaaccgggca	13380
			gcctaatgga			13440
			tcttgaaccc			13500
			agggtaacga			13560
			agaccctgct			13620
			tccgcaggcc			13680
gegetgegge	cccgcggtca	gatgctagta	gcccatttcc	aagcttgata	gggtctctta	13740
ccagcactcg	caccacccgc	ccgcgcctgc	tgggcgagga	ggagtaccta	aacaactcgc	13800
tgctgcagcc	gcagcgcgaa	aaaaacctgc	ctccggcatt	tcccaacaac	gggatagaga	13860
			cgtacgcgca			13920
			acgaccgtca			13980
			tggatttggg			14040
			tttaaaaaaa			14100
aaaaactcac	caaggccatg	gcaccgagcg	ttggttttct	tgtattcccc	ttagtatgcg	14160
gcgcgcggcg	atgtatgagg	aaggtcctcc	tccctcctac	gagagtgtgg	tgagcgcggc	14220
gccagtggcg	gcggcgctgg	gttctccctt	cgatgctccc	ctggacccgc	cgtttgtgcc	14280
tccgcggtac	ctgcggccta	ccggggggag	aaacagcatc	cgttactctg	agttggcacc	14340
			ggacaacaag			14400
gaactaccag	aacgaccaca	gcaactttct	gaccacggtc	attcaaaaca	atgactacag	14460
cccgggggag	gcaagcacac	agaccatcaa	tcttgacgac	cggtcgcact	ggggcggcga	14520
			aaatgtgaac			14580
			gcctactaag			14640
			gggcaactac			14700
			cttgaaagtg			14760
			ccgcaacttc			14820
cactggtctt	gtcatgcctg	gggtatatac	aaacgaagcc	ttccatccag	acatcattt	14880
gctgccagga	tgcggggtgg	acttcaccca	cagccgcctg	agcaacttgt	tgggcatccg	14940

caagcggcaa	cccttccagg	agggctttag	gatcacctac	gatgatctgg	agggtggtaa	15000
cattccccca	ctottggatg	tggacgccta	ccaggcgagc	ttgaaagatg	acaccgaaca	15060
gaacaaaaat	ggcgcaggcg	gcagcaacag	cagtggcagc	ggcgcggaag	agaactccaa	15120
cacaacaacc	gcggcaatgc	agccggtgga	ggacatgaac	gatcatgcca	ttcgcggcga	15180
cacctttqcc	acacgggctg	aggagaagcg	cgctgaggcc	gaagcagcgg	ccgaagctgc	15240
cacccccact	gcgcaacccg	aggtcgagaa	gcctcagaag	aaaccggtga	tcaaacccct	15300
gacagaggac	agcaagaaac	gcagttacaa	cctaataagc	aatgacagca	ccttcaccca	15360
gtaccgcagc	togtaccttg	catacaacta	cggcgaccct	cagaccggaa	tccgctcatg	15420
gaccctgctt	tocactcctg	acgtaacctg	cggctcggag	caggtctact	ggtcgttgcc	15480
agacatgatg	caagaccccg	tgaccttccg	ctccacgcgc	cagatcagca	actttccggt	15540
gatagacacc	gagctgttgc	ccgtgcactc	caagagcttc	tacaacgacc	aggccgtcta	15600
ctcccaactc	atccgccagt	ttacctctct	gacccacgtg	ttcaatcgct	ttcccgagaa	15660
ccagattttg	gcgcgccgc	cagececcae	catcaccacc	gtcagtgaaa	acgttcctgc	15720
totcacagat	cacgggacgc	taccgctgcg	caacagcatc	ggaggagtcc	agcgagtgac	15780
cattactgac	gccagacgcc	gcacctgccc	ctacgtttac	aaggccctgg	gcatagtctc	15840
accacacate	ctatcgagcc	gcactttttg	agcaagcatg	tccatcctta	tatcgcccag	15900
caataacaca	ggctggggcc	tacacttccc	aagcaagatg	tttggcgggg	ccaagaagcg	15960
ctccgaccaa	cacccagtgc	acatacacaa	gcactaccgc	gcgccctggg	gcgcgcacaa	16020
acacaaccac	actgggcgca	ccaccatca	tgacgccatc	gacgcggtgg	tggaggaggc	16080
acacaactac	acgcccacgc	caccaccagt	gtccacagtg	gacgcggcca	ttcagaccgt	16140
gatacacaa	gcccggcgct	atoctaaaat	gaagagacgg	cggaggcgcg	tagcacgtcg	16200
ccaccaccac	cgacccggca	ctgccgccca	acqcqcqqcq	geggeeetge	ttaaccgcgc	16260
acategeege	ggccgacggg	cooccatoco	ggcgctcga	aggctggccg	cgggtattgt	16320
cactataccc	cccaggtcca	adcascasac	ggccgccgca	gcagccgcgg	ccattagtgc	16380
tatgactcag	ggtcgcaggg	gcaacgtgta	ttaaatacac	gactcggtta	gcggcctgcg	16440
catacccata	cgcacccgcc	ccccccccacaa	ctagattgca	agaaaaaact	acttagactc	16500
atactattat	atgtatccag	cuacaacaac	gcgcaacgaa	gctatgtcca	agcgcaaaat	16560
gtactgttgt	atgctccagg	tratrorocc	ggagatctat	ggccccccga	agaaggaaga	16620
caaayaayay	aagccccgaa	acctagaged	ggtcaaaaag	aaaaaqaaaq	atgatgatga	16680
gcaygactac	gacgaggtgg	agetactaca	cactaccaca	cccaggcgac	gggtacagtg	16740
tgaacttgac	cgcgtaaaac	atattttaca	acccaacacc	accotactct	ttacqcccqq	16800
gaaagguuga	accegeacet	acaacccct	gtatgatgag	gtgtacggcg	acgaggacct	16860
tgagegetee	gccaacgagc	acatagegege	atttacctac	ggaaagcggc	ataaggacat	16920
gettgageag	ccgctggacg	200003333	aacacctage	ctaaagcccg	taacactgca	16980
getggegttg	cccgcgcttg	agggcaacce	ageaccuge	ggcctaaagc	gcgagtctgg	17040
gcaggtgctg	cccaccgtgc	caccycccya	agadaagege	cancoactoo	aagatgtctt	17100
tgacttggca	accgtggaac	agetgatggt	acceaagege	cacatacaac	caatcaagca	17160
ggaaaaaatg	accgtggaac	transported	gcccgaggcc	atacccacta	ccagtagcac	17220
ggtggcgccg	ggactgggcg	tgcagaccgt	ggacgcccag	tecceatta	cctcaggggt	17280
cagtattgcc	accgccacag	agggcatgga	gacacaaacg	aaracctcta	cagagataca	17340
ggcggatgcc	gcggtgcagg	eggtegetge	ggccgcgccc	ccacacaatt	cgaggaagta	17400
aacggacccg	tggatgtttc	gegetteage	teeetaest	ccttccatta	cacctaccc	17460
cggcgccgcc	agcgcgctac	tgcccgaata	cyccccacac	actacconc	accasaccac	17520
cggctatcgt	ggctacacct	accgccccay	aagacgagca	attaccegae	gccgaaccac	17580
cactggaacc	cgccgccgcc	gregeegrey	ceageeegeg	acagececya	accaccccad	17640
cagggtggct	cgcgaaggag	gcaggaccct	ggtgctgcca	acagegeget	accaccccag	17700
catcgtttaa	aagccggtct	ttgtggttct	tgcagatatg	geeetcaeet	geegeeteeg	17760
tttcccggtg	ccgggattcc	gaggaagaat	gcaccgtagg	aggggcatgg	ccggccacgg	17820
cctgacgggc	ggcatgcgtc	gtgcgcacca	ccggcggcgg	egegegeege	accordat	17880
gcgcggcggt	atcctgcccc	tccttattcc	actgatcgcc	geggegattg	gegeegtgee	17940
cggaattgca	tccgtggcct	tgcaggcgca	gagacactga	Ltaadaacaa	gttgcatgtg	18000
gaaaaatcaa	aataaaaagt	ctggactctc	acgctcgctt	ggtcctgtaa	ctattttgta	18060
gaatggaaga	catcaacttt	gcgtctctgg	ccccgcgaca	cggctcgcgc	ccgttcatgg	18120
gaaactggca	agatatcggc	accagcaata	tgagcggtgg	cgccttcagc	tggggctcgc	18120
tgtggagcgg	cattaaaaat	ttcggttcca	ccgttaagaa	ctatggcagc	aaggcctgga	18180
acagcagcac	aggccagatg	ctgagggata	agttgaaaga	gcaaaatttc	caacaaaagg	19240

tggtagatgg	cctggcctct	ggcattagcg	gggtggtgga	cctggccaac	caggcagtgc	18300
aaaataagat	taacagtaag	cttgatcccc	gccctcccgt	agaggagcct	ccaccggccg	18360
tggagacagt	gtctccagag	gggcgtggcg	aaaagcgtcc	gcgccccgac	agggaagaaa	18420
ctctggtgac	gcaaatagac	gagcctccct	cgtacgagga	ggcactaaag	caaggcctgc	18480
ccaccacccg	tcccatcgcg	cccatggcta	ccggagtgct	gggccagcac	acacccgtaa	18540
cgctggacct	gcctccccc	gccgacaccc	agcagaaacc	tgtgctgcca	ggcccgaccg	18600
ccgttgttgt	aacccgtcct	agccgcgcgt	ccctgcgccg	cgccgccagc	ggtccgcgat	18660
cgttgcggcc	cgtagccagt	ggcaactggc	aaagcacact	gaacagcatc	gtgggtctgg	18720
gggtgcaatc	cctgaagcgc	cgacgatgct	tctgaatagc	taacgtgtcg	tatgtgtgtc	18780
atgtatgcgt	ccatgtcgcc	gccagaggag	ctgctgagcc	gccgcgcgcc	cgctttccaa	18840
gatggctacc	ccttcgatga	tgccgcagtg	gtcttacatg	cacatctcgg	gccaggacgc	18900
ctcggagtac	ctgagccccg	ggctggtgca	gtttgcccgc	gccaccgaga	cgtacttcag	18960
cctgaataac	aagtttagaa	accccacggt	ggcgcctacg	cacgacgtga	ccacagaccg	19020
gtcccagcgt	ttgacgctgc	ggttcatccc	tgtggaccgt	gaggatactg	cgtactcgta	19080
caaggcgcgg	ttcaccctag	ctgtgggtga	taaccgtgtg	ctggacatgg	cttccacgta	19140
ctttgacatc	cgcggcgtgc	tggacagggg	ccctactttt	aagccctact	ctggcactgc	19200
ctacaacgcc	ctggctccca	agggtgcccc	aaatccttgc	gaatgggatg	aagctgctac	19260
tgctcttgaa	ataaacctag	aagaagagga	cgatgacaac	gaagacgaag	tagacgagca	19320
agctgagcag	caaaaaactc	acgtatttgg	gcaggcgcct	tattctggta	taaatattac	19380
aaaggagggt	attcaaatag	gtgtcgaagg	tcaaacacct	aaatatgccg	ataaaacatt	19440
tcaacctgaa	cctcaaatag	gagaatctca	gtggtacgaa	actgaaatta	atcatgcagc	19500
tgggagagtc	cttaaaaaga	ctaccccaat	gaaaccatgt	tacggttcat	atgcaaaacc	19560
cacaaatgaa	aatggagggc	aaggcattct	tgtaaagcaa	caaaatggaa	agctagaaag	19620
tcaagtggaa	atgcaatttt	tctcaactac	tgaggcgacc	gcaggcaatg	gtgataactt	19680
gactcctaaa	gtggtattgt	acagtgaaga	tgtagatata	gaaaccccag	acactcatat	19740
ttcttacatg	cccactatta	aggaaggtaa	ctcacgagaa	ctaatgggcc	aacaatctat	19800
gcccaacagg	cctaattaca	ttgcttttag	ggacaatttt	attggtctaa	tgtattacaa	19860
cagcacgggt	aatatgggtg	ttctggcggg	ccaagcatcg	cagttgaatg	ctgttgtaga	19920
tttgcaagac	agaaacacag	agctttcata	ccagcttttg	cttgattcca	ttggtgatag	19980
aaccaggtac	ttttctatgt	ggaatcaggc	tgttgacagc	tatgatccag	atgttagaat	20040
tattgaaaat	catggaactg	aagatgaact	tccaaattac	tgctttccac	tgggaggtgt	20100
gattaataca	gagactctta	ccaaggtaaa	acctaaaaca	ggtcaggaaa	atggatggga	20160
aaaagatgct	acagaatttt	cagataaaaa	tgaaataaga	gttggaaata	attttgccat	20220
ggaaatcaat	ctaaatgcca	acctgtggag	aaatttcctg	tactccaaca	tagcgctgta	20280
tttgcccgac	aagctaaagt	acagtccttc	caacgtaaaa	atttctgata	acccaaacac	20340
ctacgactac	atgaacaagc	gagtggtggc	tcccgggtta	gtggactgct	acattaacct	20400
tggagcacgc	tggtcccttg	actatatgga	caacgtcaac	ccatttaacc	accaccgcaa	20460
tgctggcctg	cgctaccgct	caatgttgct	gggcaatggt	cgctatgtgc	ccttccacat	20520
ccaggtgcct	cagaagttct	ttgccattaa	aaacctcctt	ctcctgccgg	gctcatacac	20580
ctacgagtgg	aacttcagga	aggatgttaa	catggttctg	cagagctccc	taggaaatga	20640
cctaagggtt	gacggagcca	gcattaagtt	tgatagcatt	tgcctttacg	ccaccttctt	20700
ccccatggcc	cacaacaccg	cctccacgct	tgaggccatg	cttagaaacg	acaccaacga	20760
ccagtccttt	aacgactatc	tctccgccgc	caacatgctc	taccctatac	ccgccaacgc	20820
taccaacgtg	cccatatcca	tcccctcccg	caactgggcg	gctttccgcg	gctgggcctt	20880
cacgcgcctt	aagactaagg	aaaccccatc	actgggctcg	ggctacgacc	cttattacac	20940
ctactctggc	tctataccct	acctagatgg	aaccttttac	ctcaaccaca	cctttaagaa	21000
ggtggccatt	acctttgact	cttctgtcag	ctggcctggc	aatgaccgcc	tgcttacccc	21060
caacgagttt	gaaattaagc	gctcagttga	cggggagggt	tacaacgttg	cccagtgtaa	21120
				tacaacattg		21180
cttctatatc	ccagagagct	acaaggaccg	catgtactcc	ttctttagaa	acttccagcc	21240
catgagccgt	caggtggtgg	atgatactaa	atacaaggac	taccaacagg	tgggcatcct	21300
acaccaacac	aacaactctg	gatttgttgg	ctaccttgcc	cccaccatgc	gcgaaggaca	21360
ggcctaccct	gctaacttcc	cctatccgct	tataggcaag	accgcagttg	acagcattac	21420
ccagaaaaag	tttctttgcg	atcgcaccct	ttggcgcatc	ccattctcca	gtaactttat	21480
gtccatgggc	gcactcacag	acctgggcca	aaaccttctc	tacgccaact	ccgcccacgc	21540

	acttttgagg	tagatoccat	ממפכתפתכככ	accettettt	atgttttgtt	21600
gctagacatg	gacgtggtcc	atatacacca	accacaccac	agcatcatca	aaaccgtgta	21660
tgaagtcttt	cccttctcgg	grardeacca	cacaacataa	agaagcaagc	aacatcaaca	21720
cctgcgcacg	ccatgggctc	cagtgaggg	gaactgaaag	ccattotcaa	agatettggt	21780
acagetgeeg	atttttggg	cacctatgac	aaccoctttc	caggettet	ttctccacac	21840
Egigggccat	gcgccatagt	caetacgae	aatcacaaa	ctaggaraga	acactggatg	21900
aagetegeet	ggaacccgca	ctcaaaaaca	tactacetet	ttgagcctt	tagetttet	21960
gcctttgcct	tcaagcaggt	ttaccacttt	gagtacgagt	cactcctaca	ccatagcacc	22020
gaccagcgac	cccccgaccg	stateteesa	ctagaaaaat	ccacccaaag	catacaggg	22080
attgettett	ccgcctgtgg	actattctcc	tacatattta	tccacacctt	tgccaactgg	22140
cccaactcgg	ccatggatca	actattctgc	atgaacctta	ttaccoooot	acccaactcc	22200
ccccaaactc	gtccccaggt	agagggagg	ctacatcaca	accaddagge	actotacago	22260
atgeteaaca	gccactcgcc	atacttcccc	accacacto	cacagattaa	gagcgccact	22320
ttcctggagc	acttgaaaaa	catataaaaa	taatgtacta	gagagagttt	caataaaggc	22380
tetttttgte	atttgtacac	tatagagaga	ttatttaccc	ccaccettac	catctacacc	22440
aaatgetttt	caaaggggtt	ataccacaca	teactataca	ccactagcag	ggacacgttg	22500
gtttaaaaat	gtttagtgct	ccgccgcgca	tcaccacaca	ccatcccca	cageteggtg	22560
cgatactggt	gittagiget	acceptate	accaacacat	ttagcaggt	gagagagat	22620
aagttttcac	tccacaggct	gegeaccacc	tacacacacaca	acttoccata	cacagggttg	22680
atcttgaagt	cgcagttggg	geeteegeee	tacacactaa	ccaccacact	cttatcaga	22740
cagcactgga	acactatcag	egeegggtgg	ctcacgctgg	acquartcaa	ctttaataac	22800
atcagatccg	cgtccaggtc	ctccgcgttg	tttaaattaa	acygayicaa	tagtagcatc	22860
tgccttccca	aaaagggcgc	gtgcccaggc	restagações	actogcaccy	accettate	22920
aaaaggtgac	cgtgcccggt	etgggegila	ggatacageg	cccycacaaa	agecttgace	22980
tgcttaaaag	ccacctgagc	etttgegeet	tcagagaaga	acatgccgca	agacttgccg	23040
gaaaactgat	tggccggaca	ggccgcgtcg	tgcacgcagc	terest	agagtactac	23100
atctgcacca	catttcggcc	ccaccggttc	ttcacgatct	contracts	agactgette	23160
ttcagcgcgc	gctgcccgtt	ttegetegte	acatccattt	tatesagge	cccccacc	23220
atcataatgc	ttccgtgtag	acacttaagc	tegeettega	ceteagegea	gcggcgcagc	23280
cacaacgcgc	agcccgtggg	ctcgtgatgc	ttgtaggtca	tettestest	cyactycagy	23340
tacgcctgca	ggaatcgccc	catcatcgtc	acaaaggtet	tgttgttggt	gaaggccagc	23400
tgcaacccgc	ggtgctcctc	gttcagccag	gtettgeata	eggeegeeag	agettecate	23460
tggtcaggca	gtagtttgaa	gttcgccttt	agatcgttat	ccacgtggta	ettgtctatt	23520
agcgcgcgcg	cagcctccat	gecettetee	cacgcagaca	cgatcggcac	acceageggg	23580
ttcatcaccg	taatttcact	ttccgcttcg	ctgggctctt	cetetteete	cegegeeege	23640
ataccacgcg	ccactgggtc	gtcttcattc	agccgccgca	ctgtgcgctt	accecettig	23700
ccatgcttga	ttagcaccgg	tgggttgctg	aaacccacca	tttgtagege	cacaccccc	23760
ctttcttcct	cgctgtccac	gattacctct	ggtgatggcg	ggcgctcggg	cttgggagaa	23700
gggcgcttct	ttttcttctt	gggcgcaatg	gccaaatccg	ccgccgaggt	egatggeege	23880
gggctgggtg	tgcgcggcac	cagcgcgtct	tgtgatgagt	cttcctcgtc	ctcggactcg	23940
atacgccgcc	tcatccgctt	ttttgggggc	gcccggggag	gcggcggcga	eggggaeggg	24000
gacgacacgt	cctccatggt	tgggggacgt	cgcgccgcac	egegteegeg	ctcgggggtg	24060
gtttcgcgct	gctcctcttc	ccgactggcc	atttccttct	cctataggca	gaaaaagacc	24120
atggagtcag	tcgagaagaa	ggacagccta	accgccccct	ctgagttcgc	caccaccgcc	24120
tccaccgatg	ccgccaacgc	gcctaccacc	ttccccgtcg	aggcaccccc	gcttgaggag	
gaggaagtga	ttatcgagca	ggacccaggt	tttgtaagcg	aagacgacga	ggaccgctca	24240
gtaccaacag	aggataaaaa	gcaagaccag	gacaacgcag	aggcaaacga	ggaacaagtc	24300
gggcgggggg	acgaaaggca	tggcgactac	ctagatgtgg	gagacgacgt	gctgttgaag	24360
catctgcagc	gccagtgcgc	cattatctgc	gacgcgttgc	aagagcgcag	cgatgtgccc	24420
ctcgccatag	cggatgtcag	ccttgcctac	gaacgccacc	tattctcacc	gcgcgtaccc	24480
cccaaacgcc	aagaaaacgg	cacatgcgag	cccaacccgc	gcctcaactt	ctaccccgta	24540
tttgccgtgc	cagaggtgct	tgccacctat	cacatctttt	tccaaaactg	caagataccc	24600
ctatcctgcc	gtgccaaccg	cagccgagcg	gacaagcagc	tggccttgcg	gcagggcgct	24660
gtcatacctg	atatcgcctc	gctcaacgaa	gtgccaaaaa	tctttgaggg	tcttggacgc	24720
gacgagaagc	gcgcggcaaa	cgctctgcaa	caggaaaaca	gcgaaaatga	aagtcactct	24780
ggagtgttgg	tggaactcga	gggtgacaac	gcgcgcctag	ccgtactaaa	acgcagcatc	24840

asaatasaaa	actttaccta	cccaacactt	aacctaccc	ccaacatcat	gagcacagtc	24900
						24960
					tttgcaagaa	25020
					gcttcaaacg	25080
					gctcgttacc	25140
					caagctagag	
					gatctccaac	25200
					ccttgggcaa	25260
					cgactgcgtt	25320
tacttatttc	tatgctacac	ctggcagacg	gccatgggcg	tttggcagca	gtgcttggag	25380
gagtgcaacc	tcaaggagct	gcagaaactg	ctaaagcaaa	acttgaagga	cctatggacg	25440
gccttcaacg	agcgctccgt	ggccgcgcac	ctggcggaca	tcattttccc	cgaacgcctg	25500
cttaaaaccc	tgcaacaggg	tctgccagac	ttcaccagtc	aaagcatgtt	gcagaacttt	25560
aggaacttta	tcctagagcg	ctcaggaatc	ttgcccgcca	cctgctgtgc	acttcctagc	25620
gactttgtgc	ccattaagta	ccgcgaatgc	cctccgccgc	tttggggcca	ctgctacctt	25680
			tctgacataa			25740
			ctatgcaccc			25800
					gggtccctcg	25860
			aaactcactc			25920
			cacgcccacg			25980
			gcctgcgtca			26040
			caagagtttc			26100
			ctcaacccaa			26160
			caggatggca			26220
			ctgggacagt			26280
-						26340
			ggagagccta			26400
			ctcggtcgca			26460
			aacctccgct			26520
			caccactgga			
			acagcgccaa			26580
			agactgtggg			26640
			cttcccccgt			26700
			cagcggcagc			26760
			ctctgacaaa			26820
cggcagcagc	aggaggagga	gcgctgcgtc	tggcgcccaa	cgaacccgta	tcgacccgcg	26880
			atgctatatt			26940
aacaagagct	gaaaataaaa	aacaggtctc	tgcgatccct	cacccgcagc	tgcctgtatc	27000
acaaaagcga	agatcagctt	cggcgcacgc	tggaagacgc	ggaggctctc	ttcagtaaat	27060
actgcgcgct	gactcttaag	gactagtttc	gcgccctttc	tcaaatttaa	gcgcgaaaac	27120
tacgtcatct	ccagcggcca	cacccggcgc	cagcacctgt	cgtcagcgcc	attatgagca	27180
			accagccaca			27240
			acatgagcgc			27300
gggtcaacgg	aatccgcgcc	caccgaaacc	gaattctctt	ggaacaggcg	gctattacca	27360
			gttggcccgc			27420
			gagacgccca			27480
			gtcacagggt			27540
			ttcagctcaa			27600
cacttaatet	ccatccaasc	gggacatttc	agatcggcgg	caccaaccat	ccttcattca	27660
cacctcatca	ggcaatccta	actetacada	cctcgtcctc	tgagccgcgc	tctggaggca	27720
			tgccatcggt			27780
			ttcctaactt			27840
						27900
			aggcagagca			27960
			gcgactccgg			28020
			acggcgtccg			28020
			agcgcccct			
gaccctgtgt	tctcactgtg	atttgcaact	gtcctaacct	tygattacat	caagatettt	28140

gttgccatct	ctgtgctgag	tataataaat	acagaaatta	aaatatactg	gggctcctat	28200
caccatcctg	taaacgccac	cgtcttcacc	cgcccaagca	aaccaaggcg	aaccttacct	28260
ggtactttta	acatctctcc	ctctgtgatt	tacaacagtt	tcaacccaga	cggagtgagt	28320
ctacgagaga	acctctccga	gctcagctac	tccatcagaa	aaaacaccac	cctccttacc	28380
taccaggaac	gtacgagtgc	gtcaccggcc	gctgcaccac	acctaccgcc	tgaccgtaaa	28440
ccagactttt	tccggacaga	cctcaataac	tctgtttacc	agaacaggag	gtgagcttag	28500
aaaaccctta	gogtattagg	ccaaaggcgc	agctactgtg	gggtttatga	acaattcaag	28560
caactctacq	ggctattcta	attcaggttt	ctctagaatc	ggggttgggg	ttattctctg	28620
tettataatt	ctctttattc	ttatactaac	gcttctctgc	ctaaggctcg	ccgcctgctg	28680
tatacacatt	tocatttatt	gtcagctttt	taaacgctgg	ggtcgccacc	caagatgatt	28740
aggtacataa	tcctaggttt	actcaccctt	gcgtcagccc	acggtaccac	ccaaaaggtg	28800
gattttaagg	agccagcctg	taatgttaca	ttcgcagctg	aagctaatga	gtgcaccact	28860
cttataaaat	gcaccacaga	acatgaaaag	ctgcttattc	gccacaaaaa	caaaattggc	28920
aagtatgctg	tttatgctat	ttggcagcca	ggtgacacta	cagagtataa	tgttacagtt	28980
ttccacacta	aaagtcataa	aacttttato	tatacttttc	cattttatga	aatgtgcgac	29040
attaccatgt	acatgagcaa	acagtataag	ttgtggcccc	cacaaaattg	tgtggaaaac	29100
actogcactt	tctgctgcac	tactatacta	attacagtgc	tcgctttggt	ctgtacccta	29160
ctctatatta	aatacaaaaq	cagacgcagc	tttattgagg	aaaagaaaat	gccttaattt	29220
actaagttac	aaagctaatg	tcaccactaa	ctgctttact	cgctgcttgc	aaaacaaatt	29280
caaaaagtta	gcattataat	tagaatagga	tttaaacccc	ccggtcattt	cctgctcaat	29340
accattcccc	tgaacaattg	actctatgtg	ggatatgctc	cagcgctaca	accttgaagt	29400
carrettect	ggatgtcagc	atctgacttt	ggccagcacc	tatcccgcgg	atttgttcca	29460
gtccaactac	agcgacccac	cctaacagag	atgaccaaca	caaccaacgc	ggccgccgct	29520
accogactta	catctaccac	aaatacaccc	caagtttctg	cctttgtcaa	taactgggat	29580
aacttgggca	tataataatt	ctccatageg	cttatgtttg	tatgccttat	tattatgtgg	29640
ctcatctgct	gcctaaagcg	caaacqcqcc	cgaccaccca	tctatagtcc	catcattgtg	29700
ctacacccaa	acaatgatgg	aatccataga	ttggacggac	tgaaacacat	gttcttttct	29760
cttacagtat	gattaaatga	gacatgattc	ctcgagtttt	tatattactg	accettgttg	29820
cacttttta	tacatactec	acattoocto	cggtttctca	catcgaagta	gactgcattc	29880
carcetteac	agtctatttg	ctttacggat	ttgtcaccct	cacgctcatc	tgcagcctca	29940
tcactgtggt	catcoccttt	atccagtgca	ttgactgggt	ctgtgtgcgc	tttgcatatc	30000
tcadagagaga	tececagtae	agggacagga	ctatagctga	gcttcttaga	attctttaat	30060
tatgaaattt	actgtgactt	ttctgctgat	tatttgcacc	ctatctgcgt	tttgttcccc	30120
gacctccaag	cctcaaagac	atatatcatq	cagattcact	cgtatatgga	atattccaag	30180
ttoctacaat	gaaaaaagcg	atctttccga	agcctggtta	tatgcaatca	tctctgttat	30240
agtattetae	agtaccatct	tagccctagc	tatatatccc	taccttgaca	ttggctggaa	30300
acqaataqat	gccatgaacc	acccaacttt	ccccgcgccc	gctatgcttc	cactgcaaca	30360
agttgttgcc	gacaacttta	tcccaqccaa	tcagcctcgc	cccacttctc	ccacccccac	30420
tgaaatcagc	tactttaatc	taacaggagg	agatgactga	caccctagat	ctagaaatgg	30480
acccaattat	tacagagcag	cacctactag	aaagacgcag	ggcagcggcc	gagcaacagc	30540
gcatgaatca	agagetecaa	gacatggtta	acttgcacca	gtgcaaaagg	ggtatctttt	30600
gtctggtaaa	gcaggccaaa	gtcacctacg	acagtaatac	caccggacac	cgccttagct	30660
acaagttgcc	aaccaagcgt	cagaaattgg	tggtcatggt	gggagaaaag	cccattacca	30720
taactcagca	ctcggtagaa	accgaaggct	gcattcactc	accttgtcaa	ggacctgagg	30780
atctctgcac	ccttattaag	accetataca	gtctcaaaga	tcttattccc	tttaactaat	30840
aaaaaaaaat	aataaagcat	cacttactta	aaatcagtta	gcaaatttct	gtccagttta	30900
ttcagcagca	cctccttqcc	ctcctcccag	ctctggtatt	gcagcttcct	cctggctgca	30960
aactttctcc	acaatctaaa	tggaatgtca	gtttcctcct	gttcctgtcc	atccgcaccc	31020
actatettea	tgttgttgca	gatgaagcgc	gcaagaccgt	ctgaagatac	cttcaacccc	31080
gtgtatccat	atgacacgga	aaccggtcct	ccaactgtgc	cttttcttac	tcctcccttt	31140
gtatececca	atgggtttca	agagagtccc	cctggggtac	tctctttgcg	cctatccgaa	31200
cctctagtta	cctccaatgg	catgettqcq	ctcaaaatgg	gcaacggcct	ctctctggac	31260
gaggccggca	accttacctc	ccaaaatgta	accactgtga	gcccacctct	caaaaaaacc	31320
aagtcaaaca	taaacctgga	aatatctqca	cccctcacag	ttacctcaga	agccctaact	31380
ataactacca	ccgcacctct	aatggtcgcg	ggcaacacac	tcaccatgca	atcacaggcc	31440
J-5555				-		

ccgctaaccg	tgcacgactc	caaacttagc	attgccaccc	aaggacccct	cacagtgtca	31500
gaaggaaagc	tagccctgca	aacatcaggc	cccctcacca	ccaccgatag	cagtaccctt	31560
actatcactg	cctcaccccc	tctaactact	gccactggta	gcttgggcat	tgacttgaaa	31620
gagcccattt	atacacaaaa	tggaaaacta	ggactaaagt	acggggctcc	tttgcatgta	31680
acagacgacc	taaacacttt	gaccgtagca	actggtccag	gtgtgactat	taataatact	31740
tccttgcaaa	ctaaagttac	tggagccttg	ggttttgatt	cacaaggcaa	tatgcaactt	31800
aatgtagcag	gaggactaag	gattgattct	caaaacagac	gccttatact	tgatgttagt	31860
tatccgtttg	atgctcaaaa	ccaactaaat	ctaagactag	gacagggccc	tctttttata	31920
aactcagccc	acaacttgga	tattaactac	aacaaaggcc	tttacttgtt	tacagcttca	31980
aacaattcca	aaaagcttga	ggttaaccta	agcactgcca	aggggttgat	gtttgacgct	32040
acagccatag	ccattaatgc	aggagatggg	cttgaatttg	gttcacctaa	tgcaccaaac	32100
acaaatcccc	tcaaaacaaa	aattggccat	ggcctagaat	ttgattcaaa	caaggctatg	32160
gttcctaaac	taggaactgg	ccttagtttt	gacagcacag	gtgccattac	agtaggaaac	32220
aaaaataatg	ataagctaac	tttgtggacc	acaccagete	catctcctaa	ctgtagacta	32280
aatgcagaga	aagatgctaa	actcactttg	gtcttaacaa	aatgtggcag	tcaaatactt	32340
gctacagttt	cagttttggc	tgttaaaggc	agtttggctc	caatatctgg	aacagttcaa	32400
agtgctcatc	ttattataag	atttgacgaa	aatggagtgc	tactaaacaa	tteetteetg	32460
gacccagaat	attggaactt	tagaaatgga	gatcttactg	aaggcacagc	ctatacaaac	32520
gctgttggat	ttatgcctaa	cctatcagct	tatccaaaat	ctcacggtaa	aactgccaaa	32580
agtaacattg	tcagtcaagt	ttacttaaac	ggagacaaaa	ctaaacctgt	aacactaacc	32640
attacactaa	acggtacaca	ggaaacagga	gacacaactc	caagtgcata	ctctatgtca	32700
ttttcatggg	actggtctg g	ccacaactac	attaatgaaa	tatttgccac	atcetettae	32760
actttttcat	acattgccca	agaataaaga	atcgtttgtg	ttatgtttca	acgtgtttat	32820
ttttcaattg	cagaaaattt	caagtcattt	ttcattcagt	agtatagccc	caccaccaca	32880
tagcttatac	agatcaccgt	accttaatca	aactcacaga	accctagtat	tcaacctgcc	32940
acctccctcc	caacacacag	agtacacagt	cctttctccc	cggctggcct	taaaaagcat	33000
	gtaacagaca					33060
caaacgctca	tcagtgatat	taataaactc	cccgggcagc	tcacttaagt	tcatgtcgct	33120
gtccagctgc	tgagccacag	gctgctgtcc	aacttgcggt	tgcttaacgg	gcggcgaagg	33180
agaagtccac	gcctacatgg	gggtagagtc	ataatcgtgc	atcaggatag	ggcggtggtg	33240
	gcgcgaataa					33300
ggcagtggtc	tcctcagcga	tgattcgcac	cgcccgcagc	acaaggegee	ttgteeteeg	33360 33420
	cgcaccctga					33480
aatattgttc	aaaatcccac	agtgcaaggc	getgtateca	aageteatgg	eggggaccac	33540
agaacccacg	tggccatcat	accacaagcg	caggragaci	aagtggegae	cccccataaa	33600
	ataaacatta					33660
	tgattaaaca					33720
etgecegeeg	gctatacact	gcagggaacc	gggactggaa	atattagaaa	aacacacaca	33780
	ccatggatca					33840
cacgigeata	cacttcctca tcctgaatca	ggattacaag	cacactocao	gctagaacta	gcacgtaact	33900
aacaacccac	attgtcaaag	tattacatta	addeactacad	ggaagacccc	ccagtatggt	33960
acguigue	tctgtctcaa	aaggaggtag	accatcccta	ctatacaaa	tacaccasas	34020
agegegggtt	cgtgttggtc	atagraygray	acgateceta	acaccaaca	tagtcatatt	34080
tactasagas	aaaccaggtg	caaacataac	asacadatct	acatetecaa	tctcaccact	34140
	tgtgtagtag					34200
	ttctatgtaa					34260
cagaataagg	cacacccagc	caacctacac	atteatteta	caaatcacac	acadagaaa	34320
cagaacaagc	tggaagaacc	statttttt	ttttattcca	aaagattatc	caaaacctca	34380
aaatraarat	ctattaagtg	aacqcqctcc	cctccaataa	cataatcaaa	ctctacages	34440
aaacyaayat	taatggcatt	totagogotot	tocacaatoo	cttccaaaaa	gcaaacggcc	34500
ctcacctcca	agtggacgta	aaggctaaac	ccttcaggg	gaatctcctc	tataaacatt	34560
	caaccatgcc					34620
	gaatattaag					34680
ttcagcctca	agcagcgaat	catgattgca	aaaattcagg	ttectcacag	acctgtataa	34740
	-3309000					

```
gattcaaaag cggaacatta acaaaaatac cgcgatcccg taggtccctt cgcagggcca
                                                                    34800
gctgaacata atcgtgcagg tctgcacgga ccagcgcggc cacttccccg ccaggaacct
                                                                    34860
                                                                    34920
tgacaaaaga acccacactg attatgacac gcatactcgg agctatgcta accagcgtag
ccccgatgta agctttgttg catgggcggc gatataaaat gcaaggtgct gctcaaaaaa
                                                                    34980
                                                                    35040
tcaggcaaag cctcgcgcaa aaaagaaagc acatcgtagt catgctcatg cagataaagg
caggtaagct ccggaaccac cacagaaaaa gacaccattt ttctctcaaa catgtctgcg
                                                                    35100
ggtttctgca taaacacaaa ataaaataac aaaaaaacat ttaaacatta gaagcctgtc
                                                                    35160
ttacaacagg aaaaacaacc cttataagca taagacggac tacggccatg ccggcgtgac
                                                                    35220
                                                                    35280
cgtaaaaaaa ctggtcaccg tgattaaaaa gcaccaccga cagctcctcg gtcatgtccg
gagtcataat gtaagactcg gtaaacacat caggttgatt catcggtcag tgctaaaaag
                                                                    35340
cgaccgaaat agcccggggg aatacatacc cgcaggcgta gagacaacat tacagccccc
                                                                    35400
ataggaggta taacaaaatt aataggagag aaaaacacat aaacacctga aaaaccctcc
                                                                    35460
tgcctaggca aaatagcacc ctcccgctcc agaacaacat acagcgcttc acagcggcag
                                                                    35520
cctaacagtc agccttacca gtaaaaaaga aaacctatta aaaaaacacc actcgacacg
                                                                    35580
gcaccagctc aatcagtcac agtgtaaaaa agggccaagt gcagagcgag tatatatagg
                                                                    35640
actaaaaaat gacgtaacgg ttaaagtcca caaaaaacac ccagaaaacc gcacgcgaac
                                                                    35700
ctacgcccag aaacgaaagc caaaaaaccc acaacttcct caaatcgtca cttccgtttt
                                                                    35760
cccacgttac gtaacttccc attttaagaa aactacaatt cccaacacat acaagttact
                                                                    35820
cegecetaaa acetacgtea eeegeceegt teccaegece egegecaegt cacaaactee
                                                                    35880
acccctcat tatcatattg gcttcaatcc aaaataaggt atattattga tgatg
                                                                    35935
<210> 10
<211> 5965
<212> DNA
<213> Artificial Sequence
<220>
<223> NSsuboptmut
<400> 10
gccaccatgg cccccatcac cgcctacagc cagcagacca ggggcctgct gggctgcatc
                                                                       60
                                                                      120
atcaccagcc tgaccggacg cgacaagaac caggtggagg gagaggtgca ggtggtgagc
                                                                      180
accgctaccc agagettect ggccacctgc gtgaacggcg tgtgctggac cgtgtaccac
ggageeggaa geaagaeeet ggeeggaeee aagggeeeta teaceeagat gtacaeeaat
                                                                      240
gtggatcagg atctggtggg ctggcaggcc cctcccggag ccaggagcct gacaccctgt
                                                                      300
acctgtggaa gcagcgacct gtacctggtg acacgccacg ccgatgtgat ccccgtgagg
                                                                      360
                                                                      420
cgcaggggcg attctcgcgg aagcctgctg agccctaggc ccgtgagcta cctgaagggc
ageageggag gacccetget gtgtccttct ggccatgeeg tgggcatttt tegegetgee
                                                                      480
gtgtgtacca ggggcgtggc caaagccgtg gattttgtgc ccgtggaaag catggagacc
                                                                      540
accatgegea gecetgtgtt caccgacaac ageteteece etgeegtgee ecaateatte
                                                                      600
caggtggctc acctgcacgc ccctaccgga tctggcaaga gcaccaaggt gcccgctgcc
                                                                      660
tacgccgctc agggctacaa ggtgctggtg ctgaacccca gcgtggccgc taccctgggc
                                                                      720
                                                                      780
tteggegett acatgageaa ggeecatgge ategaceeca acateegeae aggegtgege
                                                                       840
accatcacca ccggagetee cgtgacctac ageacctacg gcaagtteet ggccgatgga
                                                                      900
ggctgcagcg gaggagccta cgacatcatc atctgcgacg agtgccacag caccgacagc
accaccatcc tgggcattgg caccgtgctg gatcaggccg aaacagctgg agccaggctg
                                                                      960
gtggtgctgg ccacagctac ccctcctggc agcgtgaccg tgccccatcc caatatcgag
                                                                      1020
                                                                      1080
gaggtggccc tgagcaacac aggcgagatc cccttctacg gcaaggccat ccccatcgag
gccatccgcg gaggcaggca cctgatcttc tgccacagca agaagaagtg cgacgagctg
                                                                      1140
gctgccaagc tgagcggact gggcatcaac gccgtggcct actacagggg cctggacgtg
                                                                      1200
tcagtgatcc ccaccatcgg cgatgtggtg gtggtggcca ccgacgccct gatgacaggc
                                                                      1260
tacaccggag acttcgacag cgtgatcgac tgcaacacct gcgtgaccca gaccgtggac
                                                                      1320
ttcagcctgg accccacctt caccatcgaa accaccaccg tgcctcagga tgctgtgagc
                                                                      1380
aggagecaga ggegeggaeg caceggaagg ggeaggegeg gaatttateg etttgtgaee
                                                                      1440
cctggcgaaa ggccctctgg catgttcgac agcagcgtgc tgtgcgagtg ctacgacgct
                                                                      1500
```

ggctgcgctt	ggtacgagct	gacacccgct	gaaaccagcg	tgcgcctgcg	cgcttatctg	1560
aatacccctg	gcctgcccgt	gtgtcaggac	cacctggagt	tctgggagag	cgtgttcaca	1620
ggactgaccc	acatcgacgc	ccatttcctg	agccagacca	agcaggctgg	cgacaacttc	1680
ccctatctgg	tggcctatca	ggccaccgtg	tgtgctaggg	cccaagctcc	acctccttca	1740
tgggaccaga	tgtggaagtg	cctgatccgc	ctgaagccca	ccctgcacgg	ccctacccct	1800
ctgctgtacc	gcctgggagc	cgtgcagaac	gaggtgaccc	tgacccaccc	catcaccaag	1860
tacatcatgg	cctgcatgag	cgctgatctg	gaagtggtga	ccagcacctg	ggtgctggtg	1920
ggaggcgtgc	tggccgctct	ggctgcctac	tgcctgacca	ccggaagcgt	ggtgatcgtg	1980
ggacgcatca	tcctgagcgg	aaggcccgct	atcgtgcccg	atcgcgagtt	cctgtaccag	2040
gagttcgacg	agatggagga	gtgtgccagc	cacctgccct	acatcgagca	gggcatgcag	2100
ctggccgaac	agttcaagca	gaaggccctg	ggcctgctgc	agacagccac	caaacaggcc	2160
gaagetgeeg	ctcccgtggt	ggaaagcaag	tggagggccc	tggagacctt	ctgggctaag	2220
cacatgtgga	acttcatctc	tggcatccag	tacctggccg	gactgagcac	cctgcctggc	2280
aaccccgcta	tegecageet	gatggccttc	accgctagca	tcacctctcc	cctgaccacc	2340
cagagcaccc	tgctgttcaa	cattctgggc	ggatgggtgg	ccgctcagct	ggcccctcct	2400
tcagctgctt	ctgcctttgt	gggcgctggc	attgccggag	ccgctgtggg	cagcattggc	2460
ctgggcaaag	tgctggtgga	tattctggct	ggctatggcg	ctggcgtggc	cggagccctg	2520
gtggccttca	aggtgatgag	cggagagatg	cccagcaccg	aggacctggt	gaacctgctg	2580
cctgccattc	tgagccctgg	agccctggtg	gtgggcgtgg	tgtgtgctgc	cattctgagg	2640
cgccatgtgg	gacccggaga	gggcgctgtg	cagtggatga	accgcctgat	cgccttcgcc	2700
tctcgcggaa	accacgtgag	ccctacccac	tacgtgcctg	agagcgacgc	cgctgccagg	2760
				tgaagcgcct		2820
atcaacgagg	actgcagcac	accctgcagc	ggaagctggc	tgagggacgt	gtgggactgg	2880
atctgcaccg	tgctgaccga	cttcaagacc	tggctgcaga	gcaagctgct	gccccaactg	2940
				gcgtgtggag		3000
atcatgcaga	ccacctgtcc	ctgcggagcc	cagatcacag	gccacgtgaa	gaacggcagc	3060
atgcgcatcg	tgggccctaa	gacctgcagc	aacacctggc	acggcacctt	ccccatcaac	3120
gcctacacca	ccggaccctg	cacacccagc	cctgctccca	actacagcag	ggccctgtgg	3180
agggtggctg	ccgaggagta	cgtggaggtg	accagggtgg	gagacttcca	ctacgtgacc	3240
ggaatgacca	ccgacaacgt	gaagtgtccc	tgtcaggtgc	ccgctcccga	attttttacc	3300
gaagtggatg	gcgtgcgcct	gcatcgctat	gcccctgcct	gtaggcccct	gctgcgcgaa	3360
gaagtgacct	tccaggtggg	cctgaaccag	tacctggtgg	gcagccagct	gccctgcgag	3420
cctgagcccg	atgtggccgt	gctgaccagc	atgctgaccg	accccagcca	catcacagcc	3480
gaaaccgcta	aaaggcgcct	ggccaggggc	tctcctccaa	gcctggcctc	aagcagcgct	3540
agccagctgt	ctgctcccag	cctgaaggcc	acctgcacca	cccaccacgt	gagccccgac	3600
gccgacctga	tcgaggccaa	cctgctgtgg	cgccaggaga	tgggcggcaa	catcacccgc	3660
gtggagagcg	agaacaaggt	ggtggtgctg	gacagcttcg	accccctgcg	cgccgaggag	3720
gacgagcgcg	aggtgagcgt	gcccgccgag	atcctgcgca	agagcaagaa	gttccccgct	3780
				tgctggagag		3840
				ctcccattaa		3900
				aaagcagcgt		3960
ctggccgaac	tggccaccaa	gacctttggc	agcagcgaga	gctctgccgt	ggacagcgga	4020
acagccaccg	ctctgcctga	ccaggccagc	gacgacggcg	ataagggcag	cgatgtggag	4080
agctatagca	gcatgcctcc	cctggaaggc	gaacctggcg	atcccgatct	gagcgatggc	4140
agctggagca	ccgtgagcga	agaggccagc	gaggacgtgg	tgtgttgcag	catgagctac	4200
acctggacag	gcgctctgat	cacaccctgc	gctgccgagg	agagcaagct	gcccatcaac	4260
gccctgagca	acagcctgct	gaggcaccac	aacatggtgt	acgccaccac	cagcaggtct	4320
gccggactga	ggcagaagaa	ggtgaccttc	gaccgcctgc	aggtgctgga	cgaccactac	4380
cgcgatgtgc	tgaaggagat	gaaggccaag	gccagcaccg	tgaaggccaa	gctgctgagc	4440
gtggaggagg	cctgcaagct	gaccccccc	cacagcgcca	agagcaagtt	cggctacggc	4500
gccaaggacg	tgcgcaacct	gagcagcaag	gccgtgaacc	acatccacag	cgtgtggaag	4560
gacctgctgg	aggacaccgt	gacccccatc	gacaccacca	tcatggccaa	gaacgaggtg	4620
ttctgcgtgc	agcccgagaa	gggcggccgc	aagcccgctc	gcctgatcgt	gttccccgat	4680
ctgggcgtgc	gcgtgtgcga	gaagatggcc	ctgtacgacg	tggtgagcac	cctgcctcag	4740
gtggtgatgg	gctcaagcta	cggcttccag	tacagccctg	gccagcgcgt	ggagttcctg	4800

```
4860
gtgaacacct ggaagagcaa gaagaacccc atgggcttca gctacgacac acgctgcttc
gacagcaccg tgaccgagaa cgacatccgc gtggaggaga gcatctacca gtgctgcgac
                                                                     4920
                                                                     4980
ctggccctg aggccaggca ggccatcaag agcctgaccg agcgcctgta catcggaggc
                                                                     5040
cctctgacca acagcaaggg acagaactgc ggatacaggc gctgtagggc ctctggcgtg
ctgaccacca gctgtggcaa caccctgacc tgctacctga aggccagcgc tgcctgtcgc
                                                                     5100
gctgccaagc tgcaggactg caccatgctg gtgaacgccg ctggcctggt ggtgatttgt
                                                                     5160
gaaagcgctg gcacccagga agatgctgcc agcctgcgcg tgttcaccga ggccatgacc
                                                                     5220
aggtactctg cccctcccgg agacccccct cagcccgaat acgacctgga gctgatcacc
                                                                     5280
                                                                     5340
agctgctcaa gcaacgtgag cgtggctcac gacgccagcg gaaagcgcgt gtactacctg
acacgcgatc ccaccaccc tctggctcgc gctgcctggg aaaccgctcg ccatacaccc
                                                                     5400
                                                                     5460
gtgaacagct ggctgggcaa catcatcatg tacgccccta ccctgtgggc tcgcatgatc
                                                                     5520
ctgatgaccc acttcttcag catcctgctg gctcaggagc agctggagaa ggccctggac
                                                                     5580
tgccagattt acggcgcttg ctacagcatc gagcccctgg acctgcccca aatcatcgag
                                                                     5640
cgcctgcacg gcctgtctgc cttcagcctg cacagctaca gccctggcga aattaatcgc
gtggccagct gtctgcgcaa actgggcgtg cctcctctgc gcgtgtggag gcatagggct
                                                                     5700
aggagcgtga gggctaggct gctgagccag ggaggcaggg ccgctacctg tggaaagtac
                                                                     5760
ctgttcaact gggccgtgaa gaccaagctg aagctgaccc ctatccctgc cgctagccag
                                                                     5820
ctggacctga gcggatggtt cgtggctggc tacagcggag gcgacatcta ccacagcctg
                                                                     5880
                                                                     5940
totogogoto goootogotg gttoatgotg tgcotgctgc tgctgagcgt gggcgtgggc
                                                                     5965
atctacctgc tgcccaaccg ctaaa
<210> 11
<211> 5965
<212> DNA
<213> Artificial Sequence
<220>
<223> Chimeric NSsuboptmut
<400> 11
gccaccatgg cccccatcac cgcctacagc cagcagaccc gcggcctgct gggctgcatc
                                                                       60
                                                                      120
atcaccagcc tgaccggccg cgacaagaac caggtggagg gcgaggtgca ggtggtgagc
accgccaccc agagettect ggccacctgc gtgaacggcg tgtgctggac cgtgtaccac
                                                                      180
ggegeeggea geaagaceet ggeeggeeee aagggeeeea teacceagat gtacaceaac
                                                                      240
                                                                      300
gtggaccagg acctggtggg ctggcaggcc cccccggcg cccgcagcct gaccccctgc
                                                                      360
acctgcggca gcagcgacct gtacctggtg acccgccacg ccgacgtgat ccccgtgcgc
                                                                      420
cgccgcggcg acagccgcgg cagcctgctg agcccccgcc ccgtgagcta cctgaagggc
agcageggeg geceeetget gtgeceeage ggecaegeeg tgggeatett eegegeegee
                                                                      480
gtgtgcaccc gcggcgtggc caaggccgtg gacttcgtgc ccgtggagag catggagacc
                                                                      540
                                                                      600
accatgegea geocegtgtt caccgacaac agcageeece cegeogtgee ecagagette
caggtggccc acctgcacgc ccccaccggc agcggcaaga gcaccaaggt gcccgccgcc
                                                                      660
tacgccgccc agggctacaa ggtgctggtg ctgaacccca gcgtggccgc caccctgggc
                                                                      720
                                                                      780
tteggegeet acatgageaa ggeecaegge ategaceeca acateegeae eggegtgege
                                                                      840
accatcacca ccggcgcccc cgtgacctac agcacctacg gcaagttcct ggccgacggc
                                                                      900
ggctgcagcg gcggcgccta cgacatcatc atctgcgacg agtgccacag caccgacagc
                                                                      960
accaccatcc tgggcatcgg caccgtgctg gaccaggccg agaccgccgg cgcccgcctg
gtggtgctgg ccaccgccac ccccccggc agcgtgaccg tgccccaccc caacatcgag
                                                                      1020
                                                                      1080
gaggtggccc tgagcaacac cggcgagatc cccttctacg gcaaggccat ccccatcgag
                                                                      1140
gccatccgcg gcggccgcca cctgatcttc tgccacagca agaagaagtg cgacgagctg
geegecaage tgageggeet gggeateaac geegtggeet actacegegg eetggaegtg
                                                                      1200
agegtgatec ceaceategg egacgtggtg gtggtggeea eegacgeeet gatgacegge
                                                                      1260
tacaccggcg acttcgacag cgtgatcgac tgcaacacct gcgtgaccca gaccgtggac
                                                                      1320
ttcagcctgg accccacctt caccatcgag accaccaccg tgccccagga cgccgtgagc
                                                                      1380
cgcagccagc gccgcggccg caccggccgc ggccgccgcg gcatctaccg cttcgtgacc
                                                                      1440
cccggcgagc gccccagcgg catgttcgac agcagcgtgc tgtgcgagtg ctacgacgcc
                                                                      1500
```

						1560
				tgcgcctgcg		
				tctgggagag		1620
ggcctgaccc	acatcgacgc	ccacttcctg	agccagacca	agcaggccgg	cgacaacttc	1680
ccctacctgg	tggcctacca	ggccaccgtg	tgcgcccgcg	cccaggcccc	ccccccagc	1740
tgggaccaga	tgtggaagtg	cctgatccgc	ctgaagccca	ccctgcacgg	cccacccc	1800
ctgctgtacc	gcctgggcgc	cgtgcagaac	gaggtgaccc	tgacccaccc	catcaccaag	1860
tacatcatgg	cctgcatgag	cgccgacctg	gaggtggtga	ccagcacctg	ggtgctggtg	1920
ggcggcgtgc	tggccgccct	ggccgcctac	tgcctgacca	ccggcagcgt	ggtgatcgtg	1980
ggccgcatca	tcctgagcgg	ccgccccgcc	atcgtgcccg	accgcgagtt	cctgtaccag	2040
gagttcgacg	agatggagga	gtgcgccagc	cacctgccct	acatcgagca	gggcatgcag	2100
ctggccgagc	agttcaagca	gaaggccctg	ggcctgctgc	agaccgccac	caagcaggcc	2160
gaggccgccg	ccccgtggt	ggagagcaag	tggcgcgccc	tggagacctt	ctgggccaag	2220
cacatgtgga	acttcatcag	cggcatccag	tacctggccg	gcctgagcac	cctgcccggc	2280
aaccccgcca	tcgccagcct	gatggccttc	accgccagca	tcaccagccc	cctgaccacc	2340
				ccgcccagct		2400
agegeegeea	gcgccttcgt	gggcgccggc	atcgccggcg	ccgccgtggg	cagcatcggc	2460
ctgggcaagg	tgctggtgga	catcctggcc	ggctacggcg	ccggcgtggc	cggcgccctg	2520
gtggccttca	aggtgatgag	cggcgagatg	cccagcaccg	aggacctggt	gaacctgctg	2580
cccqccatcc	tgagccccgg	cgccctggtg	gtgggcgtgg	tgtgcgccgc	catcctgcgc	2640
caccacataa	accccaacas	agacaccata	cagtggatga	accgcctgat	cgccttcgcc	2700
agccgcggca	accacgtgag	cccacccac	tacgtgcccg	agagcgacgc	cgccgcccgc	2760
gtgacccaga	tectgageag	cctgaccatc	acccagctgc	tgaagcgcct	gcaccagtgg	2820
atcaaccaga	actocaocac	cccctgcagc	ggcagctggc	tgcgcgacgt	gtgggactgg	2880
atctgcaccg	tactaaccaa	cttcaagacc	tooctocaga	gcaagctgct	gccccagctg	2940
				gcgtgtggcg		3000
				gccacgtgaa		3060
atacacatca	taaacccaa	dacctdcadc	aacacctggc	acggcacctt	cccatcaac	3120
acgegeaceg	ccaaccccta	cacccccago	cccacccca	actacagccg	caccctataa	3180
cacataacca	ccaaccacta	catagagata	acccacataa	gcgacttcca	ctacqtqacc	3240
cacataacca	ccgaggagca	caeggaggeg	taccaaatac	ccgcccccga	gttcttcacc	3300
ggcatgacca	acatacacat	gaagagaaca	accccact	gccgccccct	gctgcgcgag	3360
gaggragacg	tecaestess	cctcaaccac	tacctootoo	gcagccagct	accetacaaa	3420
gaggtgacct	acataacat	actalceag	atactaacca	accccagcca	catcaccacc	3480
cccgagcccg	acgregatest	geegaceage	accececes	gcctggccag	caccaccacc	3540
gagaccgcca	agegeegeee	cctallage	acctocccca	cccaccacat	dadccccdac	3600
agecagetga	gegeeeeag	cetgaaggee	ccccacca	cccaccacgt	catcacccac	3660
gccgacctga	Legaggeeaa	cetgetgtgg	cyccayyaya	tgggcggcaa	caccaccege	3720
grggagageg	agaacaaggt	ggrggrgerg	atcatacaca	acccctgcg	attecesat	3780
gacgagegeg	aggregagege	geeegeegag	accetgegea	agagcaagaa	ctageagaga	3840
gccatgccca	tetgggetag	accigatiac	tatastatas	tgctggagag	acceptest	3900
cctgattacg	tgcctccagt	ggtgcatggc	cycectetyc	ctcccattaa	ageceeeee	3960
attecacete	ctaggcgcaa	aaggaccgtg	gracegacag	aaagcagcgt	gageteegee	4020
ctggccgaac	tggccaccaa	gacetttgge	agcagcgaga	gctctgccgt	ggacagegga	4020
acagccaccg	ctctgcctga	ccaggccagc	gacgacggcg	ataagggcag	cgatgtggag	4140
agctatagca	gcatgcctcc	cctggaaggc	gaacctggcg	atcccgatct	gagegatgge	
agctggagca	ccgtgagcga	agaggccagc	gaggacgtgg	tgtgttgcag	catgagetae	4200
acctggacag	gcgctctgat	cacaccctgc	gctgccgagg	agagcaagct	gcccatcaac	4260
gccctgagca	acagcctgct	gaggcaccac	aacatggtgt	acgccaccac	cagcaggtct	4320
gccggactga	ggcagaagaa	ggtgaccttc	gaccgcctgc	aggtgctgga	cgaccactac	4380
cgcgatgtgc	tgaaggagat	gaaggccaag	gccagcaccg	tgaaggccaa	gctgctgagc	4440
gtggaggagg	cctgcaagct	gaccccccc	cacagcgcca	agagcaagtt	cggctacggc	4500
gccaaggacg	tgcgcaacct	gagcagcaag	gccgtgaacc	acatccacag	cgtgtggaag	4560
gacctgctgg	aggacaccgt	gacccccatc	gacaccacca	tcatggccaa	gaacgaggtg	4620
ttctgcgtgc	agcccgagaa	gggcggccgc	aagcccgccc	gcctgatcgt	gttccccgac	4680
ctgggcgtgc	gcgtgtgcga	gaagatggcc	ctgtacgacg	tggtgagcac	cctgccccag	4740
gtggtgatgg	gcagcagcta	cggcttccag	tacagccccg	gccagcgcgt	ggagttcctg	4800

```
gtgaacacct ggaagagcaa gaagaacccc atgggcttca gctacgacac ccgctgcttc
                                                                     4860
gacagcaccg tgaccgagaa cgacatccgc gtggaggaga gcatctacca gtgctgcgac
                                                                     4920
                                                                     4980
ctggcccccg aggcccgcca ggccatcaag agcctgaccg agcgcctgta catcggcggc
cccctgacca acagcaaggg ccagaactgc ggctaccgcc gctgccgcgc cagcggcgtg
                                                                     5040
ctgaccacca gctgcggcaa caccctgacc tgctacctga aggccagcgc cgcctgccgc
                                                                     5100
gccgccaagc tgcaggactg caccatgctg gtgaacgccg ccggcctggt ggtgatctgc
                                                                     5160
gagagegeeg geacceagga ggaegeegee ageetgegeg tgttcacega ggeeatgace
                                                                     5220
cgctacagcg cccccccgg cgaccccccc cagcccgagt acgacctgga gctgatcacc
                                                                     5280
agetgeagea geaacgtgag egtggeecae gacgeeageg geaagegegt gtactacetg
                                                                     5340
                                                                     5400
accegegace ccaccacce cetggecege geegeetggg agacegeeeg ccacaccee
gtgaacagct ggctgggcaa catcatcatg tacgccccca ccctgtgggc ccgcatgatc
                                                                     5460
                                                                     5520
ctgatgaccc acttetteag cateetgetg geceaggage agetggagaa ggeeetggae
                                                                     5580
tgccagatct acggcgcctg ctacagcatc gagcccctgg acctgcccca gatcatcgag
cgcctgcacg gcctgagcgc cttcagcctg cacagctaca gccccggcga gatcaaccgc
                                                                     5640
gtggccagct gcctgcgcaa gctgggcgtg cccccctgc gcgtgtggcg ccaccgcgcc
                                                                     5700
cgcagcgtgc gcgcccgcct gctgagccag ggcggccgcg ccgccacctg cggcaagtac
                                                                     5760
ctgttcaact gggccgtgaa gaccaagctg aagctgaccc ccatccccgc cgccagccag
                                                                     5820
ctggacctga gcggctggtt cgtggccggc tacagcggcg gcgacatcta ccacagcctg
                                                                     5880
agcegegece geceegetg gtteatgetg tgeetgetge tgetgagegt gggegtggge
                                                                     5940
                                                                     5965
atctacctgc tgcccaaccg ctaaa
<210> 12
<211> 10
<212> RNA
<213> Artificial Sequence
<220>
<223> Ribosome binding site
<400> 12
                                                                        10
gccaccaugg
<210> 13
<211> 49
<212> RNA
<213> Artificial Sequence
<223> Synthetic polyadenylation signal
<400> 13
aauaaaagau cuuuauuuuc auuagaucug uguguugguu uuuugugug
                                                                        49
<210> 14
<211> 28
<212> DNA
<213> Artificial Sequence
<220>
<223> Additional nucleotides present in pVIJns-NS
                                                                        28
tctagagcgt ttaaaccctt aattaagg
<210> 15
```

<211> 15	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Additional nucleotides present in pVlJns-NSOPTmut	
<400> 15	
	15
tttaaatgtt taaac	13
<210> 16	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
•	
<220>	
<223> Oligonucleotide primer	
<400> 16	
tcgaatcgat acgcgaacct acgc	24
.010 12	
<210> 17	
<211> 37	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Oligonucleotide primer	
<400> 17	
transatate anotteana caracascan annote	37

(19) World Intellectual Property Organization International Bureau





(43) International Publication Date 17 April 2003 (17.04.2003)

(10) International Publication Number WO 03/031588 A3

C12N 15/40. (51) International Patent Classification7: 15/51, 15/85, 15/86, 15/861, A61K 48/00

I-00040 Pomezia (IT).

- (21) International Application Number: PCT/US02/32512
- (22) International Filing Date: 10 October 2002 (10.10.2002)

English

(26) Publication Language:

English

(30) Priority Data:

(25) Filing Language:

60/328,655 60/363,774

11 October 2001 (11.10.2001) US 13 March 2002 (13.03.2002)

(71) Applicants (for all designated States except US): MERCK

- & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI, S.P.A. [TT/TT]; VIA PONTINA KM. 30.600, I-00040 POMEZIA (IT).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): EMINI, Emilio, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). KASLOW, David, C. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). BETT, Andrew, J. [CA/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). NICOSIA, Alfredo [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). LAHM, Armin [DE/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). LUZZAGO, Alessandra [IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). CORTESE, Riccardo

[IT/IT]; Via Pontina KM. 30.600, I-00040 Pomezia (IT). COLLOCA, Stefano [IT/IT]; Via Pontina KM. 30.600,

- (74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GII, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, 11, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- (88) Date of publication of the international search report: 30 October 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HEPATITIS C VIRUS VACCINE

(57) Abstract: The present invention features Λd6 vectors and a nucleic acid encoding a Met-NS3-NS4Λ-NS4B-NS5Λ-NS5B polypeptide containing an inactive NSSB RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/32512

A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : C12N 15/40, 15/51, 15/85, 15/86, 15/861; A61K 48/00 US CL : 514/44; 424/93.2; 435/320.1, 455, 456 According to International Patent Classification (IPC) or to both national classification and IPC B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S.: 514/44; 424/93.2; 435/320.1, 455, 456						
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched						
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Please See Continuation Sheet						
C. DOCUMENTS CONSIDERED TO BE RELEVANT						
Category *	Citation of document, with indication, where ap	propriate,	of the relevant passages	Relevant to claim No.		
X	US 6,127,116 A (RICE et al.) 03 October 2000 (03	10.2000),	column 45, lines 18-57.	1, 2		
A	WO 01/30812 A2 (CHIRON CORPORATION) 03 May 2001 (03.05.2001).			1-54		
A	WO 97/47358 A1 (MERCK & CO., INC.) 18 December 1997 (18.12.1997).		1-54			
			See patent family annex.			
	r documents are listed in the continuation of Box C.		later document published after the inte	rnational filing date or priority		
"A" document	pecial categories of cited documents: t defining the general state of the art which is not considered to be that relevance		date and not in conflict with the applic principle or theory underlying the inve	ation but cited to understand the ation		
•	oplication or patent published on or after the international filing date	«X»	document of particular relevance; the considered novel or cannot be considered when the document is taken alone	claimed invention cannot be red to involve an inventive step		
"L" document establish specified	t which may throw doubts on priority claim(s) or which is cited to the publication date of another citation or other special reason (as	«Y»	document of particular relevance; the considered to involve an inventive step combined with one or more other such	when the document is a document, such combination		
"O" document	t referring to an oral disclosure, use, exhibition or other means		being obvious to a person skilled in th	o act		
"P" document published prior to the international filing date but later than the priority date claimed		"&" document member of the same patent family				
Date of the actual completion of the international search		Date of mailing of the international search report 0.2 SEP 2003				
09 July 2003 (09.07.2003) Name and mailing address of the ISA/US			officer O O	1 1		
Mail Stop PCT, Attn: ISA/US Commissioner for Patents		Scott D. Priebe D. Roberts 407				
P.O. Box 1450 Alexandrin, Virginia 22313-1450 Facsimile No. (703)305-3230		Telephone No. (703) 308-0196				

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/32512

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)					
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:					
1. Claim Nos.: because they relate to subject matter not required to be searched by this Authority, namely:					
Claim Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:	1				
Claim Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).					
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)					
This International Searching Authority found multiple inventions in this international application, as follows: Please See Continuation Sheet					
1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.					
 As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee. 					
As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:					
4. No required additional search fees were timely paid by the applicant. Consequently, this international search repo	rt				
is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-54					
Remark on Protest The additional search fees were accompanied by the applicant's protest. No protest accompanied the payment of additional search fees.					

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)

	PCT/US02/32512			
THE DALLETONAL OF ADOUT DEDOUT				
INTERNATIONAL SEARCH REPORT				
BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LA	CKING			
I which are not so linked as to form a single general				
inventive concept under PCT Rule 13.1. In order for all inventions to be search	ned, the appropriate additional search fees must be			
paid.				
1.				
Group I, claim(s) 1-54, drawn to a nucleic acid encoding a HCV polyprotein.				
	mence derived from human adenovirus serotypes 5 and			
Group II, claim(s) 55-59, drawn to a chimeric adenovirus vector comprising sec	Incince detined from minima agents tres serosthes a sum			
6.				
The inventions listed as Groups I and II do not relate to a single general inventi	ve concept under PCT Rule 13.1 because, under PCT			
I was a second to the terminal an approximation checkly technical resultes for the	e following reasons:			
I mt to test feeting of invention Lie a micleic scid encoding a polyb	Thielii delived hum an lic v polyprotein, whereas are			
1 to a character II is a chimeric adenoviral vector comprising a ne	derologous sequence. These two leadings are not			
related. Invention I does not require vector of invention II, nor does is the vector	or of invention II required to contain the			
polymcleotides of invention I.				
Pes,————————————————————————————————————				
Continuation of B. FIELDS SEARCHED Item 3:				
A CONTINUE DAMAGE CARTIES BIOSIS SCISEARCH, USPT. PGPB. DEL	RWENT, GENBANK, GENESEQ			
search terms: HCV, hepatitis C virus, vaccine, NSSB, NSSB near inactiv? or n	ion-functional, SEQ ID NO: 1, SEQ ID NO: 2			
1				
Ì				
•				